W. WILSON

THEORETICAL PHYSICS

2

W. WILSON

THEORETICAL PHYSICS

2

ELECTROMAGNETISM AND OPTICS
MAXWELL - LORENTZ

PREFACE

IN the present volume the exposition of the theory of physics is continued in the same spirit as in the earlier one. The chief consideration which has guided me in selecting the subject-matter has been its importance from the point of view of presenting physical theory as a coherent logical unity. Other considerations have influenced me in a minor degree and I make no claim that my choice of the material is the best that might have been made.

Among the features of this volume, to which attention may be drawn here, are : the form of the electromagnetic field equations on page 114, in which a mere change of notation brings out the 4-dimensional appearance so characteristic of the theory of relativity ; the treatment of electromagnetic momentum and mass in Chapter X, and the subsequent development, in the same chapter, of the field equations in a form which I have called an extended Poisson's equation ; the account of Huygens' principle which is based on the solution of the extended Poisson's equation, and lastly a parallel treatment of electromagnetic momentum and mass, after the manner of H. A. Lorentz, based on the FitzGerald-Lorentz contraction hypothesis. Considerable attention has been given to units and to the dimensions of physical quantities, and the electromagnetic formulae have, with some few exceptions, been developed in such a way that they are valid in *any system of units whatever*.

For convenience of reference the numbering of the sections follows on from that in Vol. I ; but, although many references are made to the earlier volume, the present one is as self-contained as it is possible for a work on electricity and optics to be, when it is remembered that these subjects constitute an organic part of the larger whole of physical science.

All the subject-matter of this volume has been taken from the notes of lectures which I have given at one time or another to university students and, while there is nothing new or original in it, the form in which it is presented has many original features, and I believe that this constitutes the main part of any value the book may have.

v

CONTENTS

CHAPTER VII

THERMODYNAMICAL ASPECTS OF ELECTRIC CURRENTS 154

CHAPTER VIII

THE ELECTROMAGNETIC THEORY OF CLERK-MAXWELL 159

CHAPTER IX

SIMPLE ELECTRON THEORY 182

THEORETICAL PHYSICS

CHAPTER I

ELECTROSTATICS

§ 18. ELECTRIFICATION

MANY materials, such as glass, ebonite, resin, etc., are observed, as a consequence of having been rubbed against one another or with other materials, to exert forces on bodies in their neighbourhood. They are said to be **electrified**. A number of glass rods, all of which have been electrified by rubbing with pieces of silk, are found to repel one another ; while the pieces of silk likewise repel one another. We cannot distinguish the state of electrification of one glass rod from that of another, nor that of one piece of silk from that of another. So we infer that bodies in the same state of electrification repel one another. But experiment shows that any one of the glass rods attracts any one of the pieces of silk. Consequently the state of electrification of the glass rods differs from that of the silk with which they were rubbed. These two states of electrification were formerly called **vitreous** and **resinous** respectively, since they were observed on glass and resin on rubbing them together. Whenever electrification is produced, both the vitreous and resinous states appear. And when the electrification is due to friction the two bodies which have been rubbed against one another always exhibit different states of electrification.

So far **electrification** is just a name for the state of a body, or of its surface, when it has acquired by friction (or, it may be, by other means) the power to exert mechanical forces on bodies in its neighbourhood ; and experiment reveals two types of electrification only.

When a brass rod is rubbed with silk or fur it does not exhibit electrification unless the precaution be taken of mounting it on a support of ebonite, glass or some other material of the class easily electrified by friction, and of holding this support in the

1   1

hand and not the piece of metal itself. This and similar facts lead us to distinguish between **conductors,** such as the piece of metal, and **insulators** or **dielectrics** like glass, ebonite, sulphur and so on. And there are of course all gradations between almost perfect insulators at one extreme and good conductors at the other.

A conductor, A, will exhibit electrification when it is merely brought into the neighbourhood of an electrified body, B. This phenomenon is called **induction,** or, more precisely, **electrostatic induction,** to distinguish it from similar phenomena to which the term ' induction ' is also applied. The end of the conductor, A, which is next to the electrified body, B, always exhibits a state of electrification unlike that on B, while the remoter end (if A is mounted on an insulating support) exhibits a state of electrification similar to that of B.

§ 18·1.   QUANTITATIVE ASPECTS OF ELECTRIFICATION

The force between two electrified particles, *not in motion* and situated in an isotropic insulating medium, is directed along the straight line joining them, and the contribution of any one of a number of electrified particles, A, B, C, . . . to the resultant force on another such particle, X, is independent of the positions and states of electrification of the remaining particles. That is to say, in calculating the force on X we have to assume that the part of it due to A, for example, is just the same as it would be if B, C, . . . were absent. These statements have an experimental basis and we shall adopt them as axioms. Furthermore, the force between two charged particles depends only on the distance between them [1] and, as we shall see, on the insulating medium in which they are situated.

We may now define the meaning of the term **quantity of electricity** or **electric charge** as follows : **The quantity of electricity on an electrified particle, A, is proportional to the force it exerts on a second electrified particle, X, the distance between them and the electrical condition of X remaining constant.** The two particles are of course assumed to be in an infinitely extended isotropic dielectric. It follows at once from the definition just given and the accompanying explanations that the force, **F,** exerted by one charged particle, A, on another, B, can be expressed in the form :

$$\mathbf{F} = e_1 e_2 \phi(r), \qquad . \quad . \quad . \quad . \quad \textbf{(18·1)}$$

[1] This is not strictly true for some insulating media which exhibit, in a faint degree, phenomena of hysteresis.

where e_1 and e_2 are the electric charges on the two particles A and B respectively, and $\phi(r)$ is the function which represents the dependence of the force $\mathbf{F}$ on the distance r between them. When the charges on the two particles are of the same kind, i.e. both of the vitreous kind or both of the resinous kind, the force will tend to separate them ; but when they are unlike it will tend to draw them together. When A and B have each the unit charge (which for the present we shall suppose to have been chosen quite arbitrarily) and are separated by the unit distance

$$\mathbf{F} = \phi(1), \quad . \quad . \quad . \quad . \quad . \quad \textbf{(18·101)}$$

we shall represent the constant $\phi(1)$ by α. Its value will, as we shall see, depend on the insulating medium surrounding the particles.

If a charged conductor be introduced into the interior of an insulated hollow conductor (the aperture in the latter being relatively very small or, better still, closed altogether after the introduction of the charged conductor by the use of a well-fitting conducting lid manipulated by an insulating thread of silk) and then caused to touch the surrounding wall it will be found, on withdrawal and testing with an electroscope, to have lost its charge completely. We shall be able to infer from this fact (§ 18·4) that the function $\phi(r)$ has the form

$$\phi(r) = \alpha/r^2. \quad . \quad . \quad . \quad . \quad . \quad \textbf{(18·11)}$$

Let us suppose the insulated hollow conductor, which we shall refer to as a Faraday vessel (Faraday's ice-pail), to be connected by a conducting wire to a gold-leaf electroscope, and a charged body to be introduced into it. The leaf of the electroscope will be observed to deflect by a definite amount, which will be quite independent of the position of the charged body in the interior of the Faraday vessel. The deflexion will also be un-affected by the introduction of other bodies (provided they are previously uncharged) and by contact between them and the original electrified body. It will be unaffected when the charged body is made to touch the surrounding Faraday vessel, in which case it will, if it is a conductor, give up the whole of its charge. In short, the deflexion is determined by the charge introduced and is quite independent of its distribution within the enclosure, of the nature of the materials in the enclosure, or of any actions or processes occurring there.

We may use the combination of electroscope and Faraday vessel to measure charges if we provide the electroscope with a scale or use a reading microscope provided with an ocular scale. We should have to calibrate the scale in some such way as the

following : We adopt some arbitrary small charge as a unit and note the deflexion produced when it is introduced into the Faraday vessel. Two bodies are now charged with this unit quantity of the same kind of electricity (i.e. both are charged with the vitreous kind of electricity for example) and introduced together into the Faraday vessel. The observed deflexion will represent two units of electricity. We then find the deflexion representing three units of electricity by introducing a body charged with two units and one charged with a single unit and so on.

Let us now suppose that two bodies *A* and *B* are charged, the former with *vitreous* and the latter with *resinous* electricity, and that the charges are found to be, for example, 5 and 3 respectively on testing them separately with the measuring device. On introducing *A* and *B* *together* into the Faraday vessel a deflexion representing 2 units will be observed, and in fact when *A* and *B* are conductors and are brought into contact with the surrounding wall of the vessel, the resulting state of electrification of the Faraday vessel and electroscope is indistinguishable from that due to 2 units of vitreous electricity. Facts such as this have led us to attach the positive sign to the one sort of electricity (vitreous) and the negative sign to the other. There is of course no compelling reason for conferring the favour of the positive sign on vitreous electricity. It might equally well have been assigned to the resinous kind.

We may now, for most purposes, drop the explicit distinction between two sorts of electricity,[1] since it will be taken care of by the sign.

When a glass rod and the silk with which it has been rubbed are introduced together into the Faraday measuring device, no deflexion is observed ; but equal deflexions when they are introduced separately. The net (algebraic) quantity generated by friction is invariably zero. This result is always obtained when charges are generated by friction. This is also true for charges produced inductively, since no change in the deflexion of the electroscope is produced by introducing into the interior of the Faraday vessel an (initially) uncharged conductor and allowing it to hang by the side of a charged body previously introduced. These and similar facts are the basis of a general law of **conservation of electric charge**, according to which we cannot

---

[1] It will force itself on our attention again when we meet with the elementary charges of electricity. The elementary positive charge is associated (in the simplest form in which it is commonly met, the proton) with a much more massive carrier than is the case with the elementary negative charge (electron). This asymmetry is one of the most remarkable facts of physical science.

alter the algebraic or net charge on a body or system of bodies without a compensating change in the charge or charges on bodies external to the system.

We have seen that when a charged conductor is made to touch the interior of a Faraday vessel it parts with its charge entirely. This means in effect that there can be no charge on the interior parts of a conductor. Charges on conductors are confined to their surfaces. This statement can be generalized as follows : **The algebraic sum of the charges within a closed conducting surface is always zero.** If for example we introduce a charged conductor into our Faraday vessel and then touch the latter, or in some way connect it conductively to the earth, the deflexion of the electroscope will drop to zero, and the Faraday vessel with its contents will behave, so far as exterior bodies are concerned, as if it were devoid of charges. If the system be insulated once again and the charged body removed, without having been in contact with the surrounding Faraday vessel, the original deflexion will be reproduced, showing the presence of the induced charge which just sufficed to make the algebraic sum of the charges in the interior equal to zero.

A closed conducting surface completely screens, as we have seen, the region exterior to it from the field within, and no change whatever that may be made in the disposition of the charges in the interior, or the character of the **electrostatic field** there, will be observable outside [1]—always provided, of course, that the change does not include the introduction or removal of charges. Experiments have also been carried out with electroscopes and other electrostatic apparatus by observers situated *within* closed conducting surfaces, or within regions which may be regarded as bounded by such surfaces, e.g. within wire cages. These have demonstrated that the conducting surface completely screens the *interior* region from an external electrostatic field.

When the *conductors* in an electrostatic field are replaced by others of different composition, the shapes of the conductors, their positions and charges remaining unchanged, the observable electrostatic phenomena are not in any way affected.

§ 18·2.　Electrostatic Fields.　Potential

An electrostatic field is completely described when we have given, for every point in it, the magnitude and direction of the

---

[1] Observable by any electrostatic measuring devices. We might indeed detect a *flow* of electricity in the interior, since one of its consequences would be a rise in temperature which would be observable *outside*.

force which would be exerted on a unit positive charge placed there, or rather, if we express it with strict accuracy, **the quotient of the force exerted on a particle with an infinitesimal charge (and finite charge density) by the amount of the charge taken with its proper sign.** This force per unit charge is termed the **field intensity (§ 2·4)** and we shall represent it by

$$\mathcal{E} \equiv (\mathcal{E}_x, \mathcal{E}_y, \mathcal{E}_z). \quad . \quad . \quad . \quad \textbf{(18·2)}$$

Attention should be paid to the way in which the definition is framed. The charged particle is to be regarded as a charged body for which the quotient, charge over volume, is finite, and the quantity which is adopted as a measure of the field intensity is the limiting value of the quotient, force over charge, when the volume of the particle becomes infinitesimal. It will be seen that, with this definition as a basis, any method adopted to measure field intensity will not do violence to the character of the field being measured.

Turning to the simple case of the field due to a single particle with the charge e, we find as a consequence of the definition of field intensity and equation **(18·1)**, that

$$\mathcal{E} = e\phi(r) \quad . \quad . \quad . \quad . \quad . \quad \textbf{(18·21)}$$

at a point distant r from the charged particle ; and if, in order to facilitate descriptions, we suppose the particle to be at the origin of rectangular co-ordinates and regard r as a vector,

$$\mathbf{r} \equiv (x, y, z), \quad . \quad . \quad . \quad . \quad \textbf{(28·22)}$$

the vector $\mathcal{E}$ will have the same direction as **(18·22)** when the charge e is positive. Therefore

$$\mathcal{E}_x/\mathcal{E} = x/\mathbf{r}, \quad \mathcal{E}_y/\mathcal{E} = y/\mathbf{r}, \quad \mathcal{E}_z/\mathcal{E} = z/\mathbf{r}.$$

Consequently

$$\left. \begin{array}{l} \mathcal{E}_x = e\phi(r)\dfrac{x}{r} = e\phi(r)\dfrac{\partial r}{\partial x}, \\[2ex] \mathcal{E}_y = e\phi(r)\dfrac{y}{r} = e\phi(r)\dfrac{\partial r}{\partial y}, \\[2ex] \mathcal{E}_z = e\phi(r)\dfrac{z}{r} = e\phi(r)\dfrac{\partial r}{\partial z}. \end{array} \right\} \quad . \quad . \quad \textbf{(18·221)}$$

Let $\mathbf{dl} \equiv (dx, dy, dz)$ represent a small displacement. The work done on a particle, with a small charge on it, when it is

displaced from the point $\mathbf{r}_0$, with the co-ordinates $(x_0,\ y_0,\ z_0)$, to another point A will be, if reckoned per unit charge,

$$V = \int_{\mathbf{r}_0}^{\mathbf{A}} (\mathbf{\mathcal{E}}\,d\mathbf{l}). \quad . \quad . \quad . \quad . \quad . \quad (18\cdot23)$$

If A be a fixed point which has been chosen (it may be quite arbitrarily) once for all, V is termed the **potential** at the point $(x_0,\ y_0,\ z_0)$. It is easy to show that it is independent of the path of integration between $\mathbf{r}_0$ and A, and that it is consequently a one-valued function of the co-ordinates $(x_0,\ y_0,\ z_0)$ of the point in question. In fact

$$V = \int_{\mathbf{r}_0}^{\mathbf{A}} \{\mathcal{E}_x dx + \mathcal{E}_y dy + \mathcal{E}_z dz\}, \quad . \quad (18\cdot231)$$

and therefore by (18·221)

$$V = \int_{\mathbf{r}_0}^{\mathbf{A}} e\phi(r)dr, \quad . \quad . \quad . \quad . \quad (18\cdot24)$$

since

$$\frac{\partial r}{\partial x}dx + \frac{\partial r}{\partial y}dy + \frac{\partial r}{\partial z}dz = dr.$$

It follows from (18·231) that

$$\mathcal{E}_x = -\frac{\partial V}{\partial x},$$

$$\mathcal{E}_y = -\frac{\partial V}{\partial y}, \quad . \quad . \quad . \quad . \quad (18\cdot25)$$

$$\mathcal{E}_z = -\frac{\partial V}{\partial z}.$$

It will be noticed that there is an arbitrary element in V, the upper limit of the defining integral being arbitrarily chosen. On the other hand, there is nothing arbitrary about the field intensity $\mathcal{E}$, which is obviously quite unaffected by our choice of the constant A. We shall often make A infinite when this is convenient, provided, of course, that the integral (28·24) converges to a finite value as A approaches infinity. In this case it will be convenient to write V in the form

$$V = \int_{r_0}^{\mathbf{A}} e\phi(r)dr = e\int_{r_0}^{\infty} \phi(r)dr = ef(r_0), . \quad . \quad (18\cdot26)$$

so that

$$\phi(r) = -\frac{df(r)}{dr}. \qquad . \qquad . \qquad . \qquad . \qquad \textbf{(18·261)}$$

It is now evident that, if we have two particles with charges e_1 and e_2, the potential due to them at a given point is, in consequence of the axioms on page 2,

$$V = e_1 f(r_1) + e_2 f(r_2), \qquad . \qquad . \qquad . \qquad \textbf{(18·262)}$$

where r_1 and r_2 are the respective distances of the particles 1 and 2 from the point in question, and more generally the potential at a given point, due to a number of point charges, will be expressed by

$$V = \underset{s}{\varSigma} e_s f(r_s). \qquad . \qquad . \qquad . \qquad . \qquad \textbf{(18·263)}$$

For a volume distribution of electricity of density, ρ, i.e. ρ units of electricity per unit volume, equation **(18·263)** may be given the form

$$V = \int \int \int \rho f(r) dx dy dz, \qquad . \qquad . \qquad . \qquad \textbf{(18·264)}$$

and when the charge is spread over one or more surfaces, instead of occupying a volume,

$$V = \int \int \sigma f(r) dS, . \qquad . \qquad . \qquad . \qquad \textbf{(18·265)}$$

σ being the surface density and dS a surface element. In both expressions the integration is extended over the whole region (volume or surface) in which there are charges.

In the formulae **(18·264)** and **(18·265)** one little difficulty has been neglected. The function $f(r)$, as we shall learn (**§ 18·4**), has the form α/r and consequently the integral **(18·264)**, for example, is made up of the elements $\alpha \rho dx dy dz/r$. If therefore ρ differs from zero at the point where V is being calculated, this element becomes indeterminate, since $\alpha \rho dx dy dz$ and r have, both of them, the limiting value zero. We shall deal with this point in **§ 18·85**.

We have seen (**§ 18·1**) that a closed conducting surface completely screens the interior region from the external field ; so that an electric field cannot exist within such a region (in static conditions) *as a consequence of an external field.* An electric field within such an enclosure can only be due to charges *within* it. In the special case where the region within the closed conducting surface is completely filled with conducting material,

charges are completely absent in the interior ; consequently the field intensity in a conductor in an electrostatic field is necessarily zero. When therefore a conductor is placed in a region where the field intensity differs from zero, a displacement of electricity (electrostatic induction, § 18) occurs in it, the ends of the conductor becoming oppositely charged. These induced charges give rise to a field which just annuls the original one in the region occupied by the conductor, so that the resultant field in the conductor is zero. All points in a conductor have therefore the same potential. We speak of this as the **potential of the conductor**. Since the potential difference between two neighbouring points, (x, y, z) and $(x + dx, y + dy, z + dz)$, is expressed by

$$- dV = \mathcal{E}_x dx + \mathcal{E}_y dy + \mathcal{E}_z dz,$$

in consequence of **(18·25)**, it is evident that in the case of two neighbouring points on the surface of a conductor

$$\mathcal{E}_x dx + \mathcal{E}_y dy + \mathcal{E}_z dz = (\mathcal{E} dl) = 0. \quad . \quad . \quad . \quad (18\cdot27)$$

This means **(§ 2·1)** *either* that $\mathcal{E}_x = \mathcal{E}_y = \mathcal{E}_z = 0$, *or* that $\mathcal{E}$ and $dl \equiv (dx, dy, dz)$ are perpendicular to one another. Therefore the electric field intensity, which of course can differ from zero *outside* the surface of the conductor, must be perpendicular to it in its immediate neighbourhood. This is merely another way of saying that the component of the intensity *parallel* to the conducting surface is zero in its immediate neighbourhood. Although therefore the field intensity in general changes discontinuously as the boundary of the conductor is crossed, its component parallel to the boundary changes continuously, being in fact zero on both sides. This is a special case of a general boundary condition which will be described later. It follows of course from **(18·27)** that *any* equipotential surface is cut orthogonally by the direction of the electric field intensity.

§ 18·3. Potential due to a Charge uniformly distributed over a Spherical Surface

Imagine a charge, Q, to be uniformly distributed over a spherical surface (Fig. 18·3), the surface density being σ, so that

$$Q = 4\pi R^2 \sigma, \quad . \quad . \quad . \quad . \quad . \quad (18\cdot3)$$

where R is the radius of the sphere.

Let OP be the straight line joining the centre of the sphere,

O, to any point P, which, in the first instance, we shall suppose
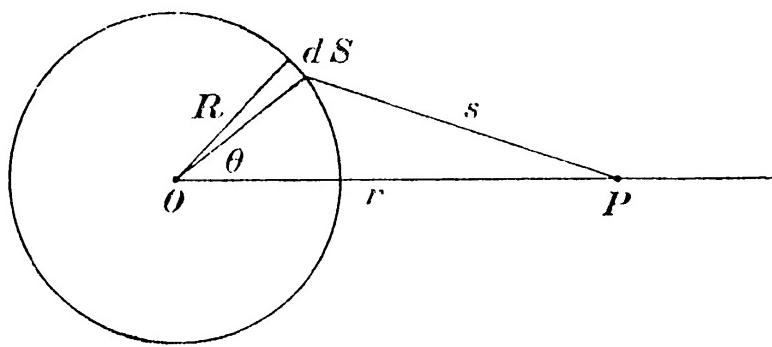
FIG. 18·3

to be outside the sphere. An element, dS, of the spherical surface, expressed in terms of polar co-ordinates, R, θ and ϕ, is

$$dS = R^2 \sin \theta \, d\theta \, d\phi,$$

and we find for the potential at P (by **18·265**)

$$V = \int_0^\pi \int_0^{2\pi} \sigma f(s) R^2 \sin \theta d\theta d\phi,$$

or

$$V = 2\pi R^2 \sigma \int_{\theta=0}^{\theta=\pi} f(s) \sin \theta d\theta. \qquad . \quad . \quad . \quad \textbf{(18·31)}$$

Now

$$s^2 = r^2 + R^2 - 2rR \cos \theta,$$

and therefore

$$2sds = 2rR \sin \theta d\theta,$$

or

$$\sin \theta d\theta = \frac{sds}{rR}.$$

On substituting this in **(18·31)** we obtain

$$V = \frac{2\pi R^2 \sigma}{rR} \int_{s=r-R}^{s=r+R} sf(s)ds,$$

or by **(18·3)**

$$V = \frac{Q}{2rR} \int_{s=r-R}^{s=r+R} sf(s)ds. \qquad . \quad . \quad . \quad \textbf{(18·311)}$$

If now we introduce a function, $F(x)$, having the property

$$\frac{dF(x)}{dx} = xf(x), \quad . \quad . \quad . \quad . \quad (18\cdot312)$$

we may express V in the form

$$V_{EXT} = \frac{Q}{2rR}\{F(r + R) - F(r - R)\}, \quad (18\cdot313)$$

where V_{EXT} is written to indicate that P is *outside* the spherical surface. If we assume P to be in the interior we must make an appropriate change in the limits of the integral in (18·311) and we thus obtain

$$V_{INT} = \frac{Q}{2rR}\{F(R + r) - F(R - r)\}. \quad . \quad (18\cdot314)$$

For a point *on* the spherical surface the two preceding expressions coalesce and give

$$V_{SUR} = \frac{Q}{2R^2}\{F(2R) - F(0)\}. \quad . \quad . \quad (18\cdot315)$$

§ 18·4. THE LAW OF FORCE

We are now in a position to deduce the **law of force** from the observational fact that the charge on a conductor is confined to its surface **(§ 18·1)**. Consider two concentric conducting surfaces, the outer one, A, with a radius a and charge Q, the inner one, B, with a radius b and charge q. The formulae **(18·313)**, **(18·314)** and **(18·315)** give us for the potentials of the two surfaces

$$V_A = \frac{Q}{2a^2}\{F(2a) - F(0)\} + \frac{q}{2ab}\{F(a + b) - F(a - b)\},$$

$$V_B = \frac{q}{2b^2}\{F(2b) - F(0)\} + \frac{Q}{2ab}\{F(a + b) - F(a - b)\}. \quad (18\cdot4)$$

Experiment shows that when the surfaces are connected by a conductor, q is zero, or so small that it cannot be detected experimentally. Therefore $q = 0$ when $V_A = V_B = V$, and equations **(18·4)** become

$$V = \frac{Q}{2a^2}\{F(2a) - F(0)\}$$

and

$$V = \frac{Q}{2ab}\{F(a + b) - F(a - b)\}.$$

Consequently

$$b\{F(2a) - F(0)\} = a\{F(a + b) - F(a - b)\}. \quad \textbf{(18·41)}$$

Now this equation will still be true if we change b while keeping a constant. Let us therefore differentiate with respect to b, a remaining constant. We get

$$\{F(2a) - F(0)\} = a\{F'(a + b) + F'(a - b)\}, \quad \textbf{(18·411)}$$

where

$$F'(x) \text{ means } \frac{dF(x)}{dx}. \quad . \quad . \quad . \quad \textbf{(18·412)}$$

On differentiating again we obtain

$$0 = a\{F''(a + b) - F''(a - b)\}, \quad . \quad . \quad \textbf{(18·413)}$$

where

$$F''(x) \text{ means } \frac{dF'(x)}{dx}. \quad . \quad . \quad . \quad \textbf{(18·414)}$$

Equation **(18·413)** simply affirms that the function $F''(x)$ is a constant. Therefore if r be the distance separating two point charges,

$$F''(r) = M \quad . \quad . \quad . \quad . \quad \textbf{(18·42)}$$

M being a constant. Consequently

$$F'(r) = Mr + N, \quad . \quad . \quad . \quad \textbf{(18·421)}$$

where N is another constant, and by **(18·312)**

$$F'(r) = rf(r).$$

Therefore

$$f(r) = M + \frac{N}{r}.\,^1$$

On differentiating this, and making use of **(18·261)**, we find

$$\phi(r) = \frac{N}{r^2}, \quad . \quad . \quad . \quad . \quad \textbf{(18·43)}$$

which expresses the law of the inverse square. Obviously N is identical with the constant α.

§ 18·5. The Experiments of Cavendish and Maxwell

The accuracy of the inverse square law was tested by Cavendish in 1773 and his experiments were repeated, with slight modifica-

---

[1] The constant M will be zero when the upper limit A of the integral **(18·24)** is infinite.

tions, by Maxwell at a much later date. All experiments of this type have failed to indicate any charge on the inner of the two conducting spherical surfaces referred to in § 18·4. Therefore $q = 0$ or is very small, and if the law of force be assumed to have the form

$$\phi(r) = \alpha r^{s-2}, \quad . \quad . \quad . \quad . \quad (18\cdot5)$$

it is evident that s, if different from zero at all, is too small to be determined. Cavendish, however, was able to estimate $|s|$, the absolute value of s, to be less than 1/50. He used a gold-leaf electroscope in his experiments. Maxwell, using the more sensitive quadrant electrometer, was able to reduce this limit to 1/21600.

Writing ϕ in the form (18·5) we have for the function f of §§ 18·3 and 18·4

$$f(r) = \int_r^\infty \phi(x)dx = \alpha \int_r^\infty x^{s-2}dx,$$

or

$$f(r) = \frac{\alpha}{1-s}r^{s-1},$$

and

$$rf(r) = \frac{\alpha r^s}{1-s}.$$

Consequently by (18·312)

$$F(r) - F(0) = \frac{\alpha r^{s+1}}{1-s^2},$$

or

$$F(r) - F(0) = \frac{\alpha r}{1-s^2}e^{s \log r}, \quad . \quad . \quad (18\cdot51)$$

and since s is known to be very small compared with unity, it will suffice to write this in the approximate form

$$F(r) - F(0) = \alpha r(1 + s \log r). \quad . \quad (18\cdot511)$$

Equations (18·4) now become

$$\left.\begin{aligned}V_A &= \frac{Q}{2a^2}A + \frac{q}{2ab}C, \\ V_B &= \frac{q}{2b^2}B + \frac{Q}{2ab}C,\end{aligned}\right\} \quad . \quad . \quad . \quad . \quad (18\cdot52)$$

where

$$A = 2\alpha a\{1 + s \log (2a)\},$$
$$B = 2\alpha b\{1 + s \log (2b)\},$$
$$C = 2\alpha b\left[1 + \frac{s}{2}\left\{\frac{a}{b} \log \frac{a+b}{a-b} + \log (a^2 - b^2)\right\}\right]$$

$$(18\cdot521)$$

Now substitute these expressions for A, B and C in $(18\cdot52)$, remembering that we may neglect not only squares and higher powers of s, but also any term involving such products as sq. We thus get

$$V_A = \frac{\alpha Q}{a}\{1 + s \log (2a)\} + \frac{\alpha q}{a},$$
$$V_B = \frac{\alpha q}{b} + \frac{\alpha Q}{a}\left[1 + \frac{s}{2}\left\{\frac{a}{b} \log \frac{a+b}{a-b} + \log (a^2 - b^2)\right\}\right]$$

$$(18\cdot522)$$

In the experimental procedure of Cavendish the two spheres, while connected by a conductor, were charged to a relatively high common potential, V. The inner one was then insulated and the outer one, consisting of separable hemispheres, removed. This was equivalent to making a infinite. The directly observed quantity was the resulting potential, V_1, of the inner sphere.[1] Cavendish was able to estimate the biggest value of V_1/V which he would just fail to detect with his electroscope, and from this estimate to compute the corresponding value of $|s|$.

It follows from the second equation $(18\cdot522)$ that

$$V_1 = \frac{\alpha q}{b},$$

and thus

$$q = \frac{bV_1}{\alpha}. \quad \quad \quad \quad \quad (18\cdot53)$$

For the original common potential, V, we have from $(18\cdot53)$ and from the second equation $(18\cdot522)$

$$V = V_1 + \frac{\alpha Q}{a}\left[1 + \frac{s}{2}\left\{\frac{a}{b} \log \frac{a+b}{a-b} + \log (a^2 - b^2)\right\}\right].$$

Whence

$$Q = \frac{a(V - V_1)}{\alpha}\left[1 - \frac{s}{2}\left\{\frac{a}{b} \log \frac{a+b}{a-b} + \log (a^2 - b^2)\right\}\right], \quad (18\cdot531)$$

---

[1] This is a convenient mode of expression. The potential, V_1, turned out in fact to be too small for measurement or detection.

while the first equation (18·522) gives us

$$V = \frac{\alpha Q}{a}\{1 + s \log (2a)\} + \frac{\alpha q}{a}. \qquad . \qquad (18\cdot532)$$

On eliminating Q and q from the three equations (18·53), (18·531) and (18·532) we get

$$\frac{a - b}{a}\frac{V_1}{V} = \frac{s}{2}\left\{\log \frac{4a^2}{a^2 - b^2} - \frac{a}{b} \log \frac{a + b}{a - b}\right\}, \qquad (18\cdot54)$$

an equation which enables the value of s to be computed.

In Maxwell's form of the experiment the outer sphere, instead of being removed, was simply earthed, which was equivalent to giving it the potential zero; i.e. $V_A = 0$. There would be a consequent very small induced charge Q_1 on the outer sphere (Q_1 vanishing of course if $s = 0$). At the same time the inner sphere would acquire a small potential V_1. Therefore we have from (18·522)

$$0 = \frac{\alpha Q_1}{a}\{1 + s \log (2a)\} + \frac{\alpha q}{a},$$

or since $Q_1 s$ is a product of two small quantities,

$$0 = Q_1 + q, \quad . \quad . \quad . \quad . \quad . \quad (18\cdot55)$$

and

$$V_1 = \frac{\alpha q}{b} + \frac{\alpha Q_1}{a}. \quad . \quad . \quad . \quad (18\cdot551)$$

If Q and V are respectively the original charge of the outer sphere and the original common potential, we have from (18·522) the further equations

$$V = \frac{\alpha Q}{a}\{1 + s \log (2a)\} + \frac{\alpha q}{a}, . \quad . \quad (18\cdot552)$$

$$V = \frac{\alpha q}{b} + \frac{\alpha Q}{a}\left[1 + \frac{s}{2}\left\{\frac{a}{b} \log \frac{a + b}{a - b} + \log (a^2 - b^2)\right\}\right]. \quad (18\cdot553)$$

On eliminating Q, q and Q_1 from the four equations (18·55), (18·551), (18·552) and (18·553) we get finally

$$\frac{V_1}{V} = \frac{s}{2}\left\{\log \frac{4a^2}{a^2 - b^2} - \frac{a}{b} \log \frac{a + b}{a - b}\right\}, \quad . \quad (18\cdot56)$$

which again enables us to compute the value of s corresponding to the estimated upper limit of the value of V_1/V.

§ 18·6.   FIELD DUE TO A DISTRIBUTION HAVING SPHERICAL
SYMMETRY

Since experiment indicates that the number s of the preceding paragraph is zero or negligibly small, we may write for the function $F(r)$, by **(18·51)** or **(18·511)**,

$$F(r) - F(0) = \alpha r. \quad . \quad . \quad . \quad \textbf{(18·6)}$$

In consequence the formulae **(18·313)**, **(18·314)** and **(18·315)** become

$$V_{EXT} = \frac{\alpha Q}{r}, \quad . \quad . \quad . \quad \textbf{(18·601)}$$

$$V_{INT} = V_{SUR} = \frac{\alpha Q}{R}. \quad . \quad . \quad \textbf{(18·602)}$$

Therefore the field intensity *outside* the sphere is

$$-\frac{\partial V}{\partial r} = \frac{\alpha Q}{r^2},$$

which is precisely what it would be if all the charge, Q, were concentrated at its centre instead of being distributed over its surface. On the other hand the potential has the same value, namely $\frac{\alpha Q}{R}$ at all points on and *within* the spherical surface. The intensity in the interior is therefore everywhere zero. We have then

$$\left. \begin{array}{l} \mathcal{E} = \dfrac{\alpha Q}{r^2} \text{ for external points,} \\[2mm] \mathcal{E} = 0 \text{ for internal points,} \end{array} \right\} \quad . \quad . \quad . \quad \textbf{(18·61)}$$

where r is the distance of the point from the centre of the sphere.

Consider now the case of two concentric spherical surfaces of radii r_1 and r_2 $(r_1 < r_2)$, each having a charge uniformly distributed over it, the respective charges being Q_1 and Q_2. At all points within *both* spheres the electric intensity will be zero ; at any point between the two spherical surfaces it will be equal to

$$\frac{\alpha Q_1}{r^2},$$

where Q_1 is the charge on the inner sphere and r is the distance of the point in question from the common centre ; because the outer sphere contributes nothing to the intensity at a point within it, while the inner one contributes as if its charge were

concentrated at its centre.   At a point distant r from the common centre, and outside *both* spheres, the intensity will be

$$\frac{\alpha(Q_1 + Q_2)}{r^2}.$$

Any spherically symmetrical distribution of charge can be regarded as built up of concentric spherical shells of infinitesimal thickness, the charge per unit area, σ, being constant over any one shell.   It appears therefore that the intensity at the distance r from the centre of symmetry is

$$\mathcal{E} = \frac{\alpha Q}{r^2}, \qquad . \quad . \quad . \quad . \quad . \quad \textbf{(18·62)}$$

where Q is the total charge *within the sphere of radius r.*

At a point within a sphere throughout which ρ is constant equation **(18·62)** gives us

$$\mathcal{E} = \frac{4\pi\alpha\rho}{3}\mathbf{r}, . \qquad . \quad . \quad . \quad . \quad \textbf{(18·63)}$$

since

$$Q = \frac{4\pi r^3 \rho}{3}.$$

If the centre of the sphere be the origin of rectangular co-ordinates $\mathbf{r}$ will be the vector

$$\mathbf{r} = (x, y, z). \qquad . \quad . \quad . \quad . \quad \textbf{(18·64)}$$

The components of $\mathcal{E}$ are consequently

$$\left.\begin{aligned}
\mathcal{E}_x &= \frac{4\pi\alpha\rho}{3}x. \\[4pt]
\mathcal{E}_y &= \frac{4\pi\alpha\rho}{3}y. \\[4pt]
\mathcal{E}_z &= \frac{4\pi\alpha\rho}{3}z.
\end{aligned}\right\} \qquad . \quad . \quad . \quad . \quad \textbf{(18·65)}$$

Hence, within the sphere,

$$\frac{\partial \mathcal{E}_x}{\partial x} + \frac{\partial \mathcal{E}_y}{\partial y} + \frac{\partial \mathcal{E}_z}{\partial z} = 4\pi\alpha\rho,$$

or $\qquad\qquad\qquad$ **div** $\mathcal{E} = 4\pi\alpha\rho.$ $\quad . \quad . \quad . \quad . \quad$ **(18·66)**

The validity of this result is not confined to the special co-ordinates we have chosen, since **(§ 2·4) div** $\mathcal{E}$ will have the same value whatever rectangular axes of co-ordinates we employ.

2

§ 18·7. The Equations of Laplace, Poisson and Gauss

Nor is the validity of **(18·66)** restricted to the special case for which it has been proved. To demonstrate this we shall first show that **div ℰ** is zero at any point in the neighbourhood of which ρ is zero, or, what amounts to the same thing, that $\nabla^2 V$ is zero **(§ 2·4)** at such a point, *whatever may be the distribution of electricity outside this neighbourhood*. The potential at any point is given by **(18·263)**, in which

$$f(r) = M + \frac{N}{r},$$

$$= \frac{\alpha}{r} \ (\S \ 18\text{·}4).$$

We have therefore

$$V = \alpha \Sigma \frac{e_s}{r_s}. \qquad \qquad \textbf{(18·7)}$$

By hypothesis there is no charge in immediate neighbourhood of the point in question and we are therefore relieved of any uncertainties which might arise from r_s in **(18·7)** becoming or approaching zero. Now it has already been shown **(§ 3·1)** that

$$\nabla^2 \frac{1}{r} = 0.$$

Therefore it follows that

$$\nabla^2 V = 0, \qquad \qquad \textbf{(18·71)}$$

or $$\textbf{div } \mathcal{E} = 0,$$

at a point in the neighbourhood of which ρ is zero. The result **(18·71)** is the **equation of Laplace** (see **§ 11**).

We now inquire about the value of **div ℰ** at a point where ρ differs from zero, and in the neighbourhood of which it varies continuously. If we imagine a sphere of radius R described with this point as centre, it is clear that, if R be made small enough, the density ρ will be sensibly constant throughout the sphere. We find therefore by **(18·66)**

$$\textbf{div } \mathcal{E} = 4\pi\alpha\rho, \qquad \qquad \textbf{(18·72)}$$

since by **(18·71)** the charges outside the sphere contribute nothing to **div ℰ**. This result is equivalent to

$$\nabla^2 V = -4\pi\alpha\rho, \qquad \qquad \textbf{(18·721)}$$

which is **Poisson's equation (§ 11)**.

Let us multiply both sides of (18·72) by the volume element $dx\,dy\,dz$ and integrate over the volume contained within a closed surface S. This gives us

$$\iiint \operatorname{div} \boldsymbol{\mathcal{E}}\, dx\,dy\,dz = 4\pi\alpha \iiint \rho\, dx\,dy\,dz.$$

Applying the theorem of Gauss (§ 3) and observing that, if Q is the total charge within the closed surface S,

$$\iiint \rho\, dx\,dy\,dz = Q,$$

we obtain the equation

$$\iint (\boldsymbol{\mathcal{E}}\,\mathbf{dS}) = 4\pi\alpha Q, \qquad . \quad . \quad . \quad . \quad \textbf{(18·73)}$$

which is also known as the theorem of Gauss.

§ 18·8. Lines and Tubes of Force—Coulomb's Law

The Gaussian formula (18·73) takes a specially simple and suggestive form when we express it in terms of Faraday's **lines of force.** These are constructed to represent the direction of the field intensity. That is to say, the tangent to a line of force at any point on it gives the direction of the field at that point. We shall introduce the convention that the number of lines passing through a surface element $\mathbf{dS}$, in the sense of its vectorial arrow, is $(\boldsymbol{\mathcal{E}},\,\mathbf{dS})$. Equation (18·73) thus becomes

$$N = 4\pi\alpha Q, \qquad . \quad . \quad . \quad . \quad . \quad \textbf{(18·8)}$$

where N is the number (in an algebraic sense) emerging through any closed surface within which the algebraic sum of the charges is Q.

It follows from (18·8) that, in an electrostatic field, the lines of force start from positive charges and end on negative charges. For when a surface encloses a region within which there are no charges at all, the algebraic sum of the lines emerging from it is zero. The lines therefore pass right through it, or else they form closed loops within it. This latter possibility is however excluded, because the line integral $\oint(\boldsymbol{\mathcal{E}},\,\mathbf{dl})$ round such a loop would differ from zero, and this is impossible in an electrostatic field by the equations (18·25) and Stokes' theorem. The lines of force must therefore pass right through any region which contains no charges. They can consequently only begin or end on electric charges. We have tacitly assumed *only one* dielectric

medium throughout the electrostatic field, or in other words we have supposed α to have everywhere the same value in the dielectric. It is possible, as we shall see, for the lines of force to begin or end at the boundary between two different dielectrics even when there is no charge there.

Obviously the lines of force cut equipotential surfaces orthogonally **(§ 18·2)**, and the definition of number of lines of force given above makes the number crossing an equipotential surface $\mathcal{E}$ per unit area.

Imagine a closed loop on an equipotential surface, and the lines of force constructed which pass through the points on the loop. The tubular region which they bound is called a **tube of force.** If ab be a small portion of the surface of a conductor enclosed within the tube of force (Fig. 18·8, in which the surface ab is perpendicular to the plane of the paper), or if the loop mentioned above bound a small area ab on the surface of a conductor, the tube will begin (or end) there. Let its cylindrical surface, indicated by ac and bd, be produced a short way into the conductor as shown by ea and fb, and suppose the resulting

FIG. 18·8

short cylinder ec, fd closed by the surfaces of arbitrary shape ef, cd, the former in the conductor and the latter in the dielectric. It is now evident that lines of force, shown in the figure by arrows, emerge (or enter) through cd only and that the number of them is

$$4\pi\alpha\sigma \times (\text{area } ab)$$

by **(18·8)**, σ being the surface density. This must be equal to

$$\mathcal{E} \times (\text{area } ab),$$

since $\mathcal{E}$ is equal to the number of lines per unit area. Consequently

$$\mathcal{E} = 4\pi\alpha\sigma, \quad . \quad . \quad . \quad . \quad \textbf{(18·81)}$$

a result which is usually termed **Coulomb's law.** This equation may of course be written

$$-\frac{\partial V}{\partial n} = 4\pi\alpha\sigma. \quad . \quad . \quad . \quad \textbf{(18·811)}$$

where $\partial V/\partial n$ represents the potential gradient just outside the conductor, n having the direction from conductor to dielectric ;
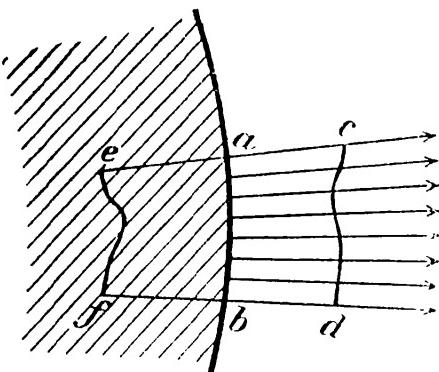
and since its value is zero within the conductor, we see that it changes discontinuously at the boundary surface of a charged conductor.

Instead of constructing lines of force we may, alternatively, make a map of an electrostatic field with the help of equipotential surfaces. Let an assemblage of these surfaces be constructed in such a way that the potential difference between every pair of consecutive surfaces has the same small value, the resulting map will represent the field intensity both as regards direction and magnitude. The intensity at any point is of course directed along the normal to the equipotential surface in which the point lies, and as its value is $- \partial V/\partial n$, its magnitude is inversely proportional to the separation of two consecutive equipotential surfaces in its neighbourhood.

§ 18·85. A Solution of Poisson's Equation

If V in Green's formula **(3·15)** be the electrostatic potential, we may replace the $\nabla^2 V$ in it by $- 4\pi\alpha\rho$ in accordance with Poisson's equation **(18·721)** and thus obtain the solution :

$$V_0 = \iiint \frac{\alpha\rho}{r}\, dx\, dy\, dz + \frac{1}{4\pi}\iint \left\{ \frac{1}{r}\frac{\partial V}{\partial n} - V\frac{\partial\left(\frac{1}{r}\right)}{\partial n} \right\} dS, \quad \textbf{(18·85)}$$

in which V_0 is the potential at any point, r is the distance from this point to the volume element $dx\, dy\, dz$ (or to dS), and the surface integral is extended over a surface bounding any region containing the point at which the potential is V_0. This formula therefore enables us to calculate the potential at any point in a region, provided we are given the values of V and $\partial V/\partial n$ at all points on its boundary, and the distribution of charge within it. It is convenient to have the co-ordinate origin at the point, O, where the potential, V_0, is to be calculated ; for then we may regard r as a vector with the components x, y and z.

Fig. 18·85

To illustrate the full significance of **(18·85)** we shall apply it to some special cases. In the first place let us suppose that all the charges are contained within a limited region, *abc* (Fig. 18·85), and that therefore the rest of space is devoid of charges. We are also assuming that the whole of space is occupied by a single dielectric medium. Let us inquire about
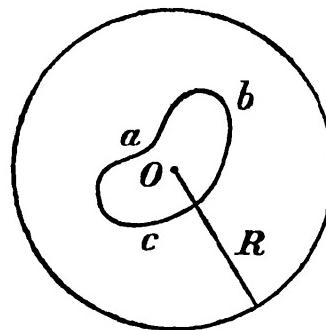
the potential, V_0, at some point, O, within, or near, abc. According to (18·264) it is given by

$$V_0 = \int\int\int \frac{\alpha\rho}{r}\, dx\, dy\, dz, \quad . \quad . \quad . \quad \textbf{(18·851)}$$

the integration extending over the whole of the region where charges exist, i.e. over the region abc or, if we chose, over any larger volume containing abc; since by hypothesis the integration over the extra volume contributes nothing to (18·851). To see that the solution (18·85) is in agreement with (18·851), let us imagine the volume integration to be extended over a sphere of very large radius, R, having its centre at O. Since R is very large, the value of V at any point on the spherical surface will be, at most, of the order of $1/R$, and therefore the terms in the brackets, { }, will be, at most, of the order $1/R^3$. On the other hand the area of the spherical surface is $4\pi R^2$, and consequently the surface integral is of the order of magnitude $1/R$, at the outside. It vanishes therefore for a sphere of infinite radius. Thus the expression for the potential reduces to (18·851), and incidentally we learn the surface integral must be equal to zero whatever R may be, provided, of course, that it is large enough to enclose the whole of abc.

The artificial character of the preceding illustration is intended to simplify it and need cause no misgivings. The following problem is much less artificial, and elucidates the meaning of the surface integral in (18·85). In Fig. 18·851 abc is a spherical surface forming the boundary between a dielectric medium, filling the interior, and a surrounding conducting medium. We imagine a charge Q to be concentrated at the centre, O', and that the surrounding conductor is maintained at some fixed potential V. The problem is to find the potential V_0 at a point O at a distance s from the centre. As in the preceding example, it can be solved by elementary means, and without invoking the aid of (18·85). We have in fact
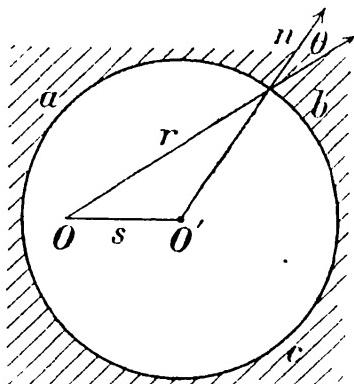
Fig. 18·851

$$V_0 = \alpha Q\{1/s - 1/R\} + V. \quad . \quad . \quad \textbf{(18·852)}$$

In applying (18·85) we may suppose the surface over which the surface integral extends, to be just within the spherical surface, separated from it by an infinitesimal interval. The volume integration then includes merely the charge Q at the

centre O' and yields the contribution $\alpha Q/s$. The surface integral consists of two parts, the first part being

$$\frac{1}{4\pi}\int\int \frac{1}{r}\frac{\partial V}{\partial n}dS. \quad . \quad . \quad . \quad (18\text{·}853)$$

Now $\partial V/\partial n$ is identical with the component of $\mathcal{E}$, the electric intensity, perpendicular to the spherical surface in the *inward* direction. By **(18·81)** we may imagine a charge density, σ, on the surface, given by

$$\frac{\partial V}{\partial n} = 4\pi\alpha\sigma.^1$$

The contribution **(18·853)** to the surface integral thus becomes

$$\int\int \frac{\alpha\sigma}{r}dS,$$

or

$$\alpha\sigma\int\int \frac{dS}{r},$$

since σ is constant in the present case. Now it can easily be shown that $\int\int \dfrac{dS}{r} = 4\pi R.^2$ This part of the surface integral is therefore

$$4\pi\alpha\sigma R,$$

or

$$\frac{4\pi\alpha\sigma R^2}{R}$$

or

$$\frac{\alpha Q'}{R}.$$

Now Q' must be equal to $-Q$. Hence we obtain

$$-\alpha Q/R. \quad . \quad . \quad . \quad (18\text{·}854)$$

The remaining part of the surface integral is

$$-\frac{1}{4\pi}\int\int V\frac{\partial\left(\dfrac{1}{r}\right)}{\partial n}dS.$$

---

¹ This does not contradict **(18·811)** since n has here the opposite direction.

² The proof of this is contained virtually in § 18·3 and is left to the reader.

In the present case V is a constant and

$$\frac{\partial\left(\frac{1}{r}\right)}{\partial n} = -\frac{1}{r^2}\frac{\partial r}{\partial n} = -\frac{1}{r^2}\cos\theta,\text{[1]}$$

where θ is the angle between the directions of r and n. The contribution to the integral is therefore

$$\frac{1}{4\pi}V\int\int\frac{\cos\theta dS}{r^2}.$$

Obviously

$$\frac{\cos\theta dS}{r^2} = d\omega,$$

where $d\omega$ is the solid angle subtended by dS at O. Therefore we have

$$\frac{V}{4\pi}\int\int d\omega = \frac{V}{4\pi}.4\pi = V. \qquad . \qquad . \qquad \textbf{(18·855)}$$

The three contributions $\alpha Q/s$, **(18·854)** and **(18·855)**, yield, as we expect, precisely the expression we have found by more elementary methods.

Had we taken the surface *outside abc*, the first part of the surface integral would have yielded zero, since in the conducting medium $\partial V/\partial n$ is zero ; the second part of the surface integral would have given us just the same expression as before, namely V ; but the volume integration in such a case would have included the charge induced on *abc* and would thus have given us both the terms $\alpha Q/s$ and $-\alpha Q/R$, with the same final result as before.

This illustration brings to light the significance of the two portions of the surface integral. We may regard the first part of it as representing the contribution to the potential at O due to charges spread over the surface and having a surface density $\partial V/4\pi\alpha\partial n$, while the other part takes account of the influence on the potential at O of the potential (or potentials) imposed on the boundary surface.

It may be pointed out here that the difficulty referred to in § **18·2** is solved by the solution **(18·85)**, or, strictly speaking, by Green's formula **(3·15)** out of which it emerges. It appears

---

[1] If, for example, r were a line drawn from the origin, and n were the X axis, $\partial r/\partial x = x/r = \cos\theta$; where θ is the angle between the directions of r and x.

from the derivation of this latter formula that if the indeterminacy in such an integral as

$$\int\int\int \frac{1}{r} \nabla^2 V \, dx \, dy \, dz,$$

which arises in consequence of $\nabla^2 V$ differing from zero at the point $r = 0$, be removed by describing a sphere with this point as centre, excluding the spherical volume from the integration, *and then* proceeding to the limit $r = 0$, we obtain the correct value for V_0. Such an integral is therefore always to be interpreted in this way (see § 3·1).

§ 18·9. Distribution of Electricity over Conducting Surfaces. Equipotential Surfaces

When an isolated charged conductor has a simple geometrical shape it is easy to determine the distribution of electricity over its surface. In the case of an isolated sphere the surface density, σ, has obviously the same value at all points on the surface. In the case of an isolated ellipsoidal conductor the value of σ at any point is proportional to the length of the perpendicular
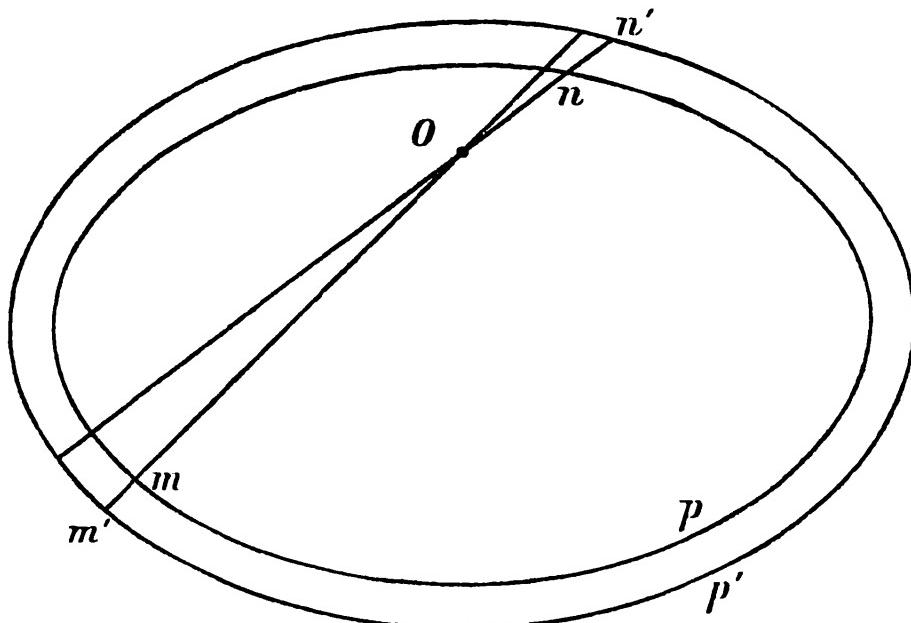
Fig. 18·9

from the centre of the ellipsoid to the tangent plane at the point in question. This can be shown in the following way :
Let the charged conducting ellipsoidal surface be represented by m, n, p (Fig. 18·9). The intensity at any point o in the interior is of course zero (§ 18·2). Imagine a cone of infinitesimal solid angle $d\omega$, with its apex at o. Let it cut the conducting

surface in the surface elements dS_m and dS_n. It is clear that σ has to conform to the condition

$$\frac{\sigma_m dS_m}{(mo)^2} = \frac{\sigma_n dS_n}{(no)^2}. \qquad . \quad . \quad . \quad . \quad \textbf{(18·9)}$$

It is convenient to suppose the section shown in Fig. 18·9 to pass through the centre of the ellipsoid. If we now construct a second surface, m', n', p', outside the ellipsoid m, n, p, and separated from it everywhere by a very small (infinitesimal) distance and imagine the interspace to be occupied by electricity of constant volume density, ρ, instead of having a surface distribution over m, n, p, we can easily show that the condition of zero field intensity at o leads to

$$(mm') = (nn'), \quad . \quad . \quad . \quad . \quad \textbf{(18·901)}$$

provided both are infinitesimal. For then we must have, instead of **(18·9)**,

$$\frac{\rho dA_m (mm')}{(mo)^2} = \frac{\rho dA_n (nn')}{(no)^2},$$

where dA_m and dA_n are the projections of dS_m and dS_n on planes perpendicular to (mn), and consequently

$$\frac{dA_m}{(mo)^2} = \frac{dA_n}{(no)^2} = d\omega.$$

The result **(18·901)** *must hold for any point, O, and for any direction (mn)*, and consequently the outer surface $m'n'p'$ must be an ellipsoid similar, and similarly situated, to the ellipsoid mnp. If dn represent the normal separation of the two surfaces at any point, the equivalent surface density, σ, is represented by

$$\sigma = \rho dn.$$

Let a, b, c be the semi-axes of the conducting surface. Those of the outer surface will be

$$a(1 + \varepsilon), \quad b(1 + \varepsilon), \quad c(1 + \varepsilon),$$

where ε is an infinitesimal number. If P be the length of the perpendicular from o to the tangent plane at any point of m, n, p, it is easy to show that

$$dn = \varepsilon P.$$

Therefore

$$\sigma = \rho dn = \varepsilon \rho P. \qquad . \quad . \quad . \quad \textbf{(18·902)}$$

The total charge, Q, on the surface, m, n, p, is equal to the product of ρ and the volume of the interspace. Hence

$$Q = \frac{4\pi abc}{3}\{(1 + \varepsilon)^3 - 1\}\rho,$$

or $$Q = 4\pi abc\varepsilon\rho,$$

since ε is infinitesimal; or finally by **(18·902)**

$$\sigma = \frac{P}{4\pi abc}\cdot Q \quad . \quad . \quad . \quad . \quad . \quad \textbf{(18·91)}$$

A simple verification of this result is obtained by applying it to the case of a sphere. Here the perpendicular P and the semi-axes a, b and c are all equal to the radius r of the sphere, and so **(18·91)** becomes

$$\sigma = \frac{Q}{4\pi r^2},$$

as we should expect.

If a suitable system of reference be used, the co-ordinates (x, y, z) of any point on the ellipsoidal surface satisfy the equation

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1. \quad . \quad . \quad . \quad . \quad \textbf{(18·92)}$$

Consider now a point (x, y, z) *outside* this surface. Obviously (x, y, z) will satisfy the equation

$$\frac{x^2}{a^2 + \lambda} + \frac{y^2}{b^2 + \lambda} + \frac{z^2}{c^2 + \lambda} = 1, \quad . \quad \textbf{(18·921)}$$

provided a suitable value be given to λ. Given x, y and z the equation **(18·921)** is a cubic equation in λ and if $a^2 > b^2 > c^2$ the three values of λ are all different. This can be made evident with the help of Fig. 18·91, which represents the relation between $f(\lambda)$ and λ, where $f(\lambda)$ means the left-hand member of equation **(18·921)**. There is some positive real value, λ_1, of λ which satisfies **(18·921)** and therefore gives $f(\lambda)$ the value unity; for any point outside the ellipsoid **(18·92)** can be regarded as situated on a confocal ellipsoid the semi-axes of which exceed a, b and c respectively. As λ diminishes from λ_1, the function $f(\lambda)$ increases, approaching $+ \infty$ as λ approaches $- c^2$. For values of λ between $- c^2$ and $- b^2$, $f(\lambda)$ varies from $- \infty$ to $+ \infty$, and therefore there is a value, λ_2, between $- c^2$ and $- b^2$ for which $f(\lambda) = 1$. Similarly there is a third value, λ_3, between $- b^2$ and $- a^2$ which
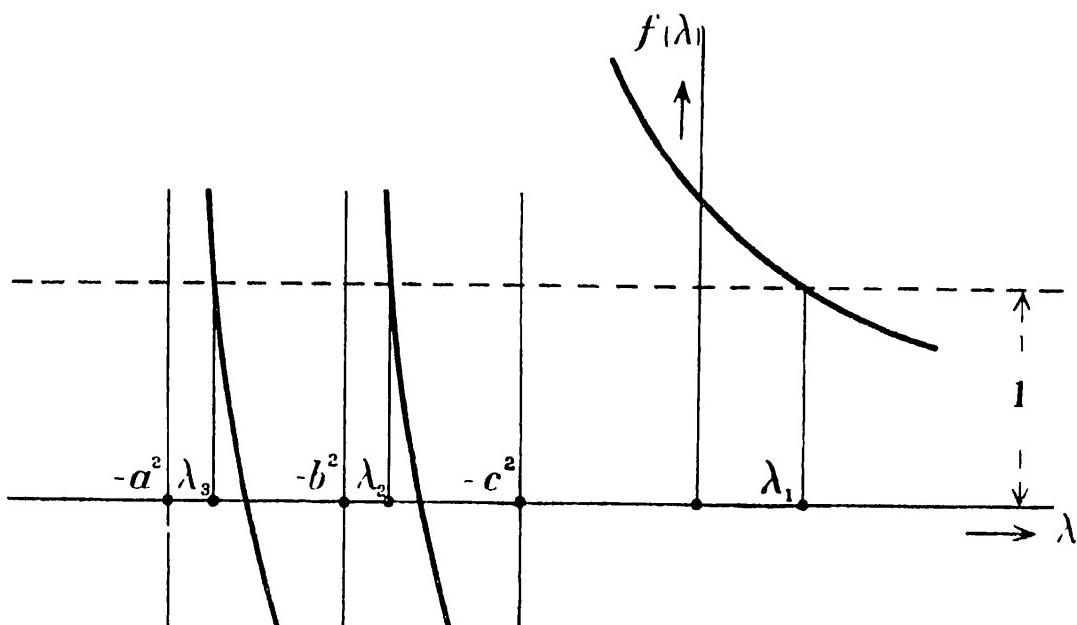
Fig. 18·91

satisfies $f(\lambda) = 1$. The point (x, y, z) is therefore the point of intersection of the three surfaces

$$\left.\begin{aligned} \frac{x^2}{a^2 + \lambda_1} + \frac{y^2}{b^2 + \lambda_1} + \frac{z^2}{c^2 + \lambda_1} &= 1, \\ \frac{x^2}{a^2 + \lambda_2} + \frac{y^2}{b^2 + \lambda_2} + \frac{z^2}{c^2 + \lambda_2} &= 1, \\ \frac{x^2}{a^2 + \lambda_3} + \frac{y^2}{b^2 + \lambda_3} + \frac{z^2}{c^2 + \lambda_3} &= 1, \end{aligned}\right\} \qquad . \quad \textbf{(18·93)}$$

of which one is an ellipsoid and the latter two of which are hyperboloids. If we subtract the first of the equations **(18·93)** from the second one we get

$$(\lambda_1 - \lambda_2)\left\{\frac{x^2}{(a^2+\lambda_1)(a^2+\lambda_2)} + \frac{y^2}{(b^2+\lambda_1)(b^2+\lambda_2)} + \frac{z^2}{(c^2+\lambda_1)(c^2+\lambda_2)}\right\} = 0,$$

and since $\lambda_1 \neq \lambda_2$,

$$\frac{x^2}{(a^2+\lambda_1)(a^2+\lambda_2)} + \frac{y^2}{(b^2+\lambda_1)(b^2+\lambda_2)} + \frac{z^2}{(c^2+\lambda_1)(c^2+\lambda_2)} = 0.$$

$$\textbf{(18·94)}$$

If D_1 and D_2 be defined by the positive square roots of the expressions

$$D_1{}^2 = \frac{x^2}{(a^2 + \lambda_1)^2} + \frac{y^2}{(b^2 + \lambda_1)^2} + \frac{z^2}{(c^2 + \lambda_1)^2},$$

$$D_2{}^2 = \frac{x^2}{(a^2 + \lambda_2)^2} + \frac{y^2}{(b^2 + \lambda_2)^2} + \frac{z^2}{(c^2 + \lambda_2)^2},$$

the direction cosines of the normals to the two surfaces at (x, y, z) are

$$(\alpha_1, \beta_1, \gamma_1) \equiv \left(\frac{x}{a^2 + \lambda_1}, \frac{y}{b^2 + \lambda_1}, \frac{z}{c^2 + \lambda_1}\right) \bigg/ D_1$$

and

$$(\alpha_2, \beta_2, \gamma_2) \equiv \left(\frac{x}{a^2 + \lambda_2}, \frac{y}{b^2 + \lambda_2}, \frac{z}{c^2 + \lambda_2}\right) \bigg/ D_2$$

respectively, and equation **(18·94)** is equivalent to

$$\alpha_1\alpha_2 + \beta_1\beta_2 + \gamma_1\gamma_2 = 0. \quad . \quad . \quad \textbf{(18·941)}$$

Therefore the two normals are at right angles to one another. It is obvious that a similar result is true for any pair of the three surfaces. Therefore all three intersect in the point (x, y, z) *orthogonally*. Instead of using the rectangular co-ordinates (x, y, z) of the point, we may for many purposes more conveniently use $(\lambda_1, \lambda_2, \lambda_3)$, which we may term its elliptical coordinates. All points on and outside the ellipsoidal conducting surface **(18·92)** will be represented by values of λ_1, λ_2 and λ_3 ranging from 0 to ∞, $-c^2$ to $-b^2$ and $-b^2$ to $-a^2$ respectively. The intersections of the associated hyperboloids constitute curves which cut the ellipsoidal surfaces including **(18·92)** orthogonally. These curves are therefore the lines of force, while the ellipsoidal surfaces are associated equipotential surfaces. In the special case where, for example, $b^2 = c^2$ and the ellipsoid **(18·92)** becomes a figure of revolution about an axis, it is easily seen that the lines of force are the intersections of a set of hyperboloids and a set of planes intersecting in the axis of revolution.

Many problems concerned with the distribution of electricity on spherical surfaces and other electrostatical questions may be solved by the method of images devised by Lord Kelvin (see § 20·4 and the sections immediately following it), or by the methods of inversion and conformal representation by means of which we can, when we have solved one problem of equipotential surfaces and lines of force, obtain the solution of another.

CHAPTER II

DISPLACEMENT AND POLARIZATION—STRESS AND ENERGY IN DIELECTRIC MEDIA

§ 19. ELECTRICAL CAPACITY

IMAGINE a conducting surface (E in Fig. 19) enclosing a region which, except for the space occupied by one or more conductors, is completely filled by a dielectric. We may conveniently assign the potential zero to this surface. Any charge given to one of the conductors, A, will, as we know, be associated with induced charges on the other conductors, and on the interior of the surrounding surface ; the algebraic sum of all these charges being zero. The space occupied by the dielectric will constitute an electrostatic field with a definite field intensity at every point. We may imagine the whole of this space mapped out by equipotential surfaces (among which will be included those of the conductors) and by lines of force cutting them orthogonally. If an exactly similar configuration of charges, i.e. one for which the surface density, σ, has precisely the same value at the same points, were superposed on the initial one, the resulting field intensity at every point in the dielectric would remain unchanged in direction ; but would have its absolute value doubled (in consequence of the axiom at the beginning of § 18·1). Suppose now that only *one* conductor, A, within the region enclosed by E has had a charge communicated to it, all other charges being induced charges consequent on the charging of A. It is clear that doubling the charge on A will have the consequence that the sign of σ and the direction of the field intensity, ℰ, will remain everywhere unaffected, but their
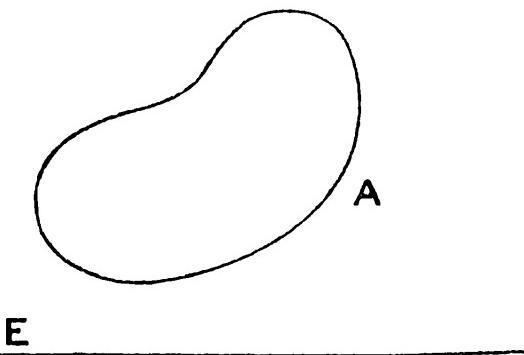
FIG. 19

absolute values will be doubled. A very obvious extension of this argument enables us to assert that the **multiplication of the charge on A by any factor, a, leaves the sign of σ and the direction of the field intensity, ε, everywhere unchanged, but multiplies their absolute values by a.** If therefore the charge on A be equal to Q,

$$\mathcal{E} = \mathbf{b}Q, \quad . \quad . \quad . \quad . \quad . \quad . \quad (19)$$

where $\mathcal{E}$ is the field intensity at a given point, and $\mathbf{b}$ is a vector which is independent of Q and is a function of (x, y, z) the co-ordinates of the point in question. In fact $\mathbf{b}$ is the electric intensity at this point associated with the unit charge on A. The potential, V, of the conductor is expressed by

$$V = \int_{A}^{E} (\mathcal{E}\,d\mathbf{l}), \quad . \quad . \quad . \quad . \quad (19\cdot01)$$

the path of integration starting at any point on the conductor, A, and ending on the surrounding surface E. It is convenient to suppose it to coincide with a line of force, though this of course is not necessary. We have then

$$V = Q \int_{A}^{E} (\mathbf{b}\,d\mathbf{l}). \quad . \quad . \quad . \quad (19\cdot011)$$

The integral

$$\int_{A}^{E} (\mathbf{b}\,d\mathbf{l}) \quad . \quad . \quad . \quad . \quad . \quad (19\cdot02)$$

is a constant, for given positions and shapes of the conductors in the region, and for a given dielectric medium. It is in fact the potential of the conductor A when its charge is the unit one. The *reciprocal* of **(19·02)** is termed the **capacity** of A. If we represent it by C, we have from **(19·011)**

$$Q = CV. \quad . \quad . \quad . \quad . \quad . \quad (19\cdot03)$$

When A is a spherical conductor, and when the surrounding conducting surface is likewise spherical and concentric with A, the intervening space being wholly filled with the dielectric medium, the integral **(19·02)** becomes

$$\frac{1}{C} = \alpha \int_{R}^{r} \frac{dr}{r^2},$$

or

$$\frac{1}{C} = \alpha \left\{ \frac{1}{R} - \frac{1}{r} \right\},$$

where R is the radius of the sphere and r that of the surrounding conducting surface. Hence

$$C = \frac{Rr}{\alpha(r - R)}. \quad . \quad . \quad . \quad . \quad \textbf{(19·04)}$$

It will be seen that the capacity of such an arrangement of conductors can be increased indefinitely by allowing r to approach indefinitely close to R in value. Any arrangement of this type is called a **condenser**—this special example being a spherical condenser.

Imagine the radius, r, of the surrounding conducting surface increased, so that $1/r$ becomes negligible by comparison with $1/R$. The capacity **(19·04)** then becomes

$$C = \frac{R}{\alpha}. \quad . \quad . \quad . \quad . \quad \textbf{(19·041)}$$

This is the expression for the capacity of an isolated spherical conductor.

Let the conductor, A, be a very long cylinder, so that we may assume its length to be infinite by comparison with its radius, R. Suppose further that the surrounding conducting surface is a co-axial cylinder of radius r, $r > R$. If the cylinder, A, be charged, the charge per unit length in regions far removed from its ends will be sensibly constant. Let it be Q. The lines of force will pass radially from the inner cylinder, A, to the surrounding one, the number intersecting the unit length of any co-axial cylinder of radius r' being

$$4\pi\alpha Q$$

by **(18·8)**. The number cutting the unit area of it will therefore be

$$\frac{4\pi\alpha Q}{2\pi r'} \quad \text{or} \quad \frac{2\alpha Q}{r'}.$$

Hence the integral **(19·02)** is in this case

$$2\alpha \int_{R}^{r} \frac{dr'}{r'} = 2\alpha \log \frac{r}{R},$$

and consequently

$$C = \frac{1}{2\alpha \log \frac{r}{R}}, \quad . \quad . \quad . \quad . \quad \textbf{(19·05)}$$

where C is the capacity per unit length.

Let us represent $r - R$ in **(19·04)** and **(19·05)** by d and suppose $d \ll R$. The two formulae then become

$$C = \frac{R^2}{\alpha d} \qquad \qquad \text{(19·06)}$$

and

$$C = \frac{R}{2\alpha d}, \qquad \qquad \text{(19·061)}$$

respectively, the former being the capacity of the spherical condenser, the latter that of the unit length of the cylindrical condenser.[1] The area of the spherical condenser is $4\pi R^2$. That of the unit length of the cylindrical one is $2\pi R$; and therefore we find for the capacity per unit area, in both cases, the expression

$$\frac{1}{\pi\alpha d}, \qquad \qquad \text{(19·062)}$$

and for the capacity of some restricted area S the expression

$$C = \frac{S}{4\pi\alpha d}. \qquad \qquad \text{(19·07)}$$

This is the expression for the capacity of a condenser consisting of two parallel plates; since such a condenser may be regarded as part of a spherical condenser of enormous radius. The lines of force, over such an area, S, will run perpendicularly to the two conducting surfaces; and the number per unit area cutting any surface perpendicular to them, i.e. parallel to the conducting surfaces, will be everywhere the same. If however we were to cut the portion, S, out and separate it altogether, this would no longer be true and the formula **(19·07)** would then have approximate validity only. In order to secure the accuracy of **(19·07)** for a parallel plate condenser Lord Kelvin introduced the device of the **guard ring**, illustrated in Fig. 20·2 (B). The effective part of the condenser is represented by the plate cd of area S and the corresponding portion of the larger plate, ef, opposite to it. The guard ring, g, surrounds cd and is separated from it by a very narrow insulating annular space. The effect of this arrangement is that, when a potential difference is established between cd and ef (the guard ring, g, having the same potential as cd) the electric intensity in the region between cd and ef will be everywhere perpendicular to the two plates

---

[1] Since $\log \frac{r}{R} = \log\left(1 + \frac{r - R}{R}\right) = \log\left(1 + \frac{d}{R}\right)$; and as d/R is very small, $\log\left(1 + \frac{d}{R}\right) = \frac{d}{R}$.

and uniform, and the capacity of this portion will be correctly given by **(19·07)**. A similar device is also employed with cylindrical condensers.

It is easy to compare the capacities of condensers experimentally. This can be done by comparing directly the charges associated with equal differences of potential between the plates of the condenser. Such experiments (first carried out by Cavendish and somewhat later by Faraday) reveal that **the capacity is not in any way dependent on the material of the conductors ; but that it does depend on the nature of the dielectric medium.** For instance, two spherical condensers, for which R and r have the same respective values, are found to differ in capacity when the dielectrics in their interspaces are different. **This means that the constant α has different values for different insulating media.** It is usual to call the *reciprocal* of α the dielectric constant of the medium. In order, however, to meet the difficulties due to the multiplicity of absolute electric units in common use, we shall write

$$\alpha = \frac{A}{K}, \quad . \quad . \quad . \quad . \quad . \quad \textbf{(19·08)}$$

where A is a pure number and merely introduced to facilitate the passage from one system of units to another. We shall usually assign to it one or other of the values

$$\left. \begin{array}{l} A = 1, \\ \text{or} \quad A = 1/4\pi \end{array} \right\} \quad . \quad . \quad . \quad . \quad \textbf{(19·081)}$$

K is the **dielectric constant.** We may therefore write for the force between two charged particles (**§ 18·4**)

$$\mathbf{F} = \frac{A}{K}\frac{e_1 e_2}{r^2}, \quad . \quad . \quad . \quad . \quad \textbf{(19·082)}$$

while the formulae **(19·04)** *et seq.* for capacities become

$$C = \frac{KrR}{A(r - R)} \text{ (spherical condenser)}, \quad . \quad . \quad . \quad \textbf{(19·09)}$$

$$C = \frac{KR}{A} \text{ (isolated sphere)}, . \quad . \quad . \quad . \quad . \quad \textbf{(19·091)}$$

$$C = \frac{K}{2A \log \dfrac{r}{R}} \text{ (unit length of cylindrical condenser)}, \textbf{(19·092)}$$

$$C = \frac{KS}{4\pi A d} \text{ (parallel plate condenser)}. \quad . \quad . \quad \textbf{(19·093)}$$

As a rule A is given the value unity. When the value $1/4\pi$

is adopted for A we shall call the units **Lorentz-Heaviside Units.** Having decided what value shall be assigned to the numerical constant, A, we may completely define our system of units (for electrostatic phenomena) by assigning the value unity to the dielectric constant, K, of some arbitrarily chosen dielectric medium. When the medium so chosen is empty space—if we may be permitted to use the term ' medium ' for empty space—we shall term the units **electrostatic units.**

§ 19·1. PHYSICAL UNITS—DIMENSIONS OF PHYSICAL QUANTITIES

At the outset of this work (**§ 2**) we considered a particular physical quantity, which we called a **displacement.** When we abstract from its vectorial character, and give our attention solely to its absolute value, we speak of it as a **length.** It is one of the simplest physical quantities, and one of those by means of which we endeavour to express other physical quantities. The unit used to measure it is fixed arbitrarily. For scientific purposes the **centimetre** has been adopted ; but we might equally well (apart from considerations of convenience) have chosen any one of a multitude of other units. The next in order of the physical quantities we have met is **time.** This is likewise measured by an arbitrary unit, the **mean solar second.** It is true that in a certain sense these units are not arbitrary. The centimetre, for instance, is determined by a larger unit, the **metre,** which was originally intended to be equal to the distance from the equator to the pole multiplied by 10^{-7}. The mean solar second has a simple numerical relationship to the mean period of the rotation of the earth relatively to the sun. These facts however do not fix the precise magnitude of the centimetre and the second. The metre, for example, has turned out to differ appreciably from 10^{-7} times the quadrant ; but its one-hundredth part continues to be our scientific unit of length. Most physical units, as we shall see, are, or can be, *precisely* determined in terms of the values adopted in *ad hoc* fashion for the units of length, time and for certain other physical quantities. Such units are said to be **derived** from the arbitrarily chosen, or **fundamental, units.** As an instance of a derived unit we may take that of **velocity** (or **speed** as we shall term it when we abstract from its vectorial character). We may of course, and we sometimes do, measure speeds in terms of arbitrary units ; but we nearly always determine them by the formula

$$v = \frac{dx}{dt}, \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad (19 \cdot 1)$$

where v is the speed and dx is the short distance travelled by the particle or other object in the short time dt, or in some equivalent way. And it is clear that, when we have fixed the units of length and time, that of speed is precisely determined by **(19·1)**. We might of course use such a formula as

$$v = k\frac{dx}{dt}, \qquad . \quad . \quad . \quad . \quad . \quad (19\cdot11)$$

where k is a constant arbitrarily assigned *once for all*. This constant does not measure or express any property of a particular body or medium, or any physical quantity or relationship. It is a mere number associated with the definition **(19·11)**. It has a universal significance and we shall say that it has no **dimensions.** Apart from the universal number, k, a speed is determined by the ratio of two physical quantities (or two things which we measure), namely displacement (length) and time of displacement. We describe this relationship by saying that its dimensions are those of length divided by time, or

$$LT^{-1}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (19\cdot12)$$

Another instance of a quantity usually expressed in terms of a derived unit is that of **volume.** Occasionally volumes are expressed in terms of arbitrary units such as the *pint, gallon,* etc. The **derived unit** of volume universally used is that of a cube, each side of which is equal to the arbitrarily chosen unit of length. With such a derived unit the volume of a sphere, for example, is expressed by

$$V = \frac{4}{3}\pi r^3,$$

where r is the radius of the sphere. We might of course have *defined* this volume by

$$V = k\frac{4}{3}\pi r^3,$$

where again k is an arbitrarily assigned number—not necessarily equal to k in **(19·11)**—fixed once for all, for the measurement of volumes. With this definition of volume, that of a rectangular block would be

$$V = kabc,$$

where a, b and c are its length, breadth and height respectively. The dimensions of a volume are, obviously, described by L^3.

We now turn to **force** and **mass.** In § 5 we might have defined the measure of a force by

$$F = km\,a, \qquad . \quad . \quad . \quad . \quad (19\cdot13)$$

instead of equation (5), k being any arbitrarily assigned number. In this equation we have three new measured quantities, namely **mass, force** and **acceleration.** The last named of these is usually measured by a unit derived from those of length and time. Its dimensions are obviously

$$LT^{-2},$$

since it can be measured by rate of change of speed. It is clear that we cannot express either force or mass in terms of units derived from those of length and time only, since we have only *the one equation* (19·13) or (5), connecting them both with a quantity which can be expressed in this way. We must therefore assign an arbitrary unit to one of them. Let us suppose it to be the unit of mass. The unit of force is then fixed by the defining equation (19·13), and the dimensions of force are seen to be

$$F = MLT^{-2}. \quad . \quad . \quad . \quad . \quad (19\cdot14)$$

The scientific unit of mass is the **gram.**

The **density** of a material is usually defined to be mass per unit volume. Its dimensions are

$$ML^{-3}.$$

Here again, of course, we might define density by introducing an arbitrary number k, so that

$$\text{Density} = k\frac{\text{Mass}}{\text{Volume}}.$$

The distinction sometimes made between **density** and **specific gravity** is misleading. The latter term is usually employed for density when an arbitrary unit (for example, the density of water or some other selected medium) is used. In all the examples given above the numerical factor k has been chosen to be unity, and so long as we confine our attention to mechanics, only three units need be arbitrary, all others being derivable from one or more of these three arbitrary or **fundamental units** as they are termed. As we have seen, the fundamental units chosen for scientific purposes are the **centimetre, gram** and **mean solar second.** These units and those derived from them constitute a system of units known as the **C.G.S. system.** If we replace the centimetre and gram by the **foot** and **pound** respectively, we obtain the **foot pound second system** of units. Either system of units, or any other system consisting of arbitrarily chosen fundamental units and units derived from them, is often described as **absolute** ; and a quantity measured

in terms of them is said to be expressed in **absolute units.** It is a very troublesome process to determine certain quantities, for example electrical resistance, current or electromotive force, in absolute units and consequently certain standards have been adopted on the basis of careful measurements, carried out with the purpose of making them as nearly as possible equal to the derived units for these quantities, or equal to simple numerical multiples of them. These standards have been fixed by international agreement, and are called **international units.** We have, for example, the **international ohm,** the **international ampere,** and so on.

It was generally believed in the days when the basis of Physics was purely mechanical, in the old-fashioned sense,[1] that all physical units could be derived from the units of length, mass and time, and many electrical units for example are still described as C.G.S. units. As a matter of fact, however, certain thermal units cannot be derived from one or more of the units of length, mass and time ; nor can *any* of the electric and magnetic units. To provide a sufficient basis for a comprehensive system of units it appears that we need **five** fundamental units, which are independent in the sense that none of them can be derived from one or more of the others. Turning to the province of heat, we find that heat itself is a form of energy, and can therefore be measured by **work.** It has consequently the dimensions

$$\text{Force} \times \text{Length},$$

or $$\text{ML}^2\text{T}^{-2} . \quad . \quad . \quad . \quad . \quad . \quad (19 \cdot 15)$$

Other thermal quantities, e.g. **temperature, specific heat** and **entropy,** never appear *singly* in equations containing length, mass and time or quantities expressible in terms of these. We are forced in consequence to introduce an arbitrary thermal unit. We may take it to be the unit of temperature difference, for example the centigrade degree on Kelvin's scale, which has a simple relationship to the temperature interval between the temperature of ice and water in equilibrium under normal pressure and that of water and its vapour in equilibrium under this pressure.[2] All other thermal units can now be derived from one or more of those of **length, mass, time** and **temperature.**

A further fundamental unit is needed for the domain of electricity and magnetism. Take equation **(19·082)** for example,

---

[1] See the quotation from Schuster's *Optics* in the preface to Vol. I.
[2] See §§ **12** and **15·15.**

which expresses the force exerted by one point charge on another,

$$\mathbf{F} = \frac{A}{K}\frac{e_1 e_2}{r^2}.$$

This equation contains *two new quantities*, **dielectric constant** and **charge,** along with others, namely force and length, the units of which we have already derived from those of length, mass and time. A is a numerical constant exactly like k in **(19·13).** Not only in equation **(19·082),** but in all equations containing *electric* or *magnetic* quantities, we shall find *two* of them present *in addition* to those quantities which are derivable from length, mass and time. We are therefore forced to introduce an electric (or magnetic) fundamental unit. This may be, for example, the unit of dielectric constant, which in the electrostatic system (or systems) is chosen to be that of empty space.

The dimensional equation emerging from **(19·082)** is

Force $= [\text{charge}]^2 \times [\text{length}]^{-2} \times [\text{dielectric constant}]^{-1}$,

or $\qquad\qquad\qquad$ $MLT^{-2} = E^2 L^{-2} K^{-1}$,

the dimensions of a charge are consequently expressed by

$$E = K^{1/2} M^{1/2} L^{3/2} T^{-1}. \quad . \quad . \quad . \quad \textbf{(19·16)}$$

We shall return to the subject of units and dimensional equations in **§ 24·4.** Meanwhile, we might point out that we are not compelled to adopt as fundamental units those of length, mass, time, temperature and dielectric constant. **We may choose any five quantities the unit of no one of which is derivable from the others.** For example, as equation **(19·14)** shows, we might choose a unit of force instead of mass, or **(19·16)** a unit of charge instead of one of dielectric constant. The five fundamental units *might*, for example, be those of **Volume, Force, Velocity, Entropy, Electric Charge ;** though this would not be a convenient selection.

It should be noted that all our formulae and equations have been developed in such a way that they are valid whatever selection of fundamental units may be made—if we except the cases where the use of specified units is expressly intended.

§ 19·2. Maxwell's Displacement Hypothesis

For the purposes of this section we shall make use, provisionally, of a mechanical picture of the electrostatic field. The picture is not an adequate one, and is only introduced to facilitate

the description of one aspect of such fields. Any region containing conductors and insulators we shall picture as occupied by an elastic solid—we may think of india-rubber, for example—which contains cavities. Large cavities, such as *abc*, or *def*, in Fig. 19·2, will simulate conductors ; whereas dielectric media will be simulated by the medium surrounding *abc*, where the cavities are too small, we shall suppose, to be perceptible or recognizable. Electricity may be simulated by an incompressible fluid of great density, which we imagine to fill all the cavities. It will be helpful also to think of the septa or walls of the cavities as very thin, so that the total volume is practically identical with that of the fluid itself ; and it will not require too great an imaginative effort to conceive of the total absence of gravity. Each
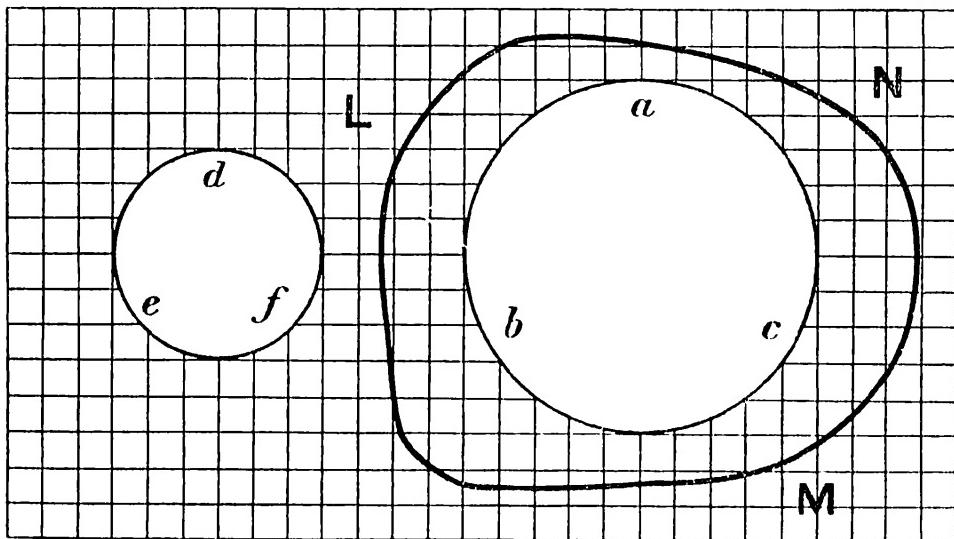
FIG. 19·2

cavity will contain a certain normal quantity of fluid, when the elastic septa are in an absolutely unstrained condition. This is the state of affairs which corresponds to the absence of charges and fields. If now we suppose the quantity of fluid in the cavity *abc* to be increased in some way, it will simulate for us a conductor in a positively charged state, the **excess of fluid** in *abc* representing the charge on the conductor *abc*. Similarly, if some of the fluid were withdrawn from *abc* a negatively charged conductor would be simulated. Charging a conductor *abc* positively will, of course, be associated with a slight *outward* movement of the bounding surface *abc*, while a negative charge will be associated with an *inward* movement. The *pressure* inside *abc* represents the *potential* of the conductor ; and the pressure at any other point represents the potential at that point, since pressure difference is equivalent to the work done (per unit volume) in transferring fluid from one point to another.

If we were to accept this mechanical model as a complete or exact representation of the state of affairs in an electrostatic field, we should of course very soon meet with insuperable difficulties ; but it does represent correctly certain important features of such a field. To begin with, we notice that when fluid is introduced into *abc*, so that it contains a quantity in excess of the normal amount (positive charge), the part of the boundary of a neighbouring cavity, *def*, nearest to *abc* will be forced inwards into *def* ; that is to say, the part of the conductor *def* on the side next to *abc* will acquire a negative charge. The remoter part of the boundary of *def* will be forced outwards. It will represent a positive charge. Moreover, the total quantity of fluid in *def* will remain unchanged, which means that the algebraic sum of the induced charges on *def* is zero.

The most significant thing, however, which our model or picture brings out, is that in the case of any closed surface, fixed in position and not moving with the medium, LMN (Fig. 19·2), enclosing the cavity (or conductor) *abc*, for example, whatever quantity of fluid may be introduced into the region within LMN, whether into the dielectric part of it or into the conducting parts such as *abc*, an equal quantity will pass outwards through LMN. More precisely the algebraic sum of the quantities of fluid introduced into the region within LMN—or created or destroyed within it—will be equal to that of the fluid which flows outwards through LMN, and this will be associated with a corresponding displacement, **D**, of the fluid at every point in the region or field. If we measure the incompressible fluid by its volume, the statement just given may be put in the form :

$$Q = \iint (\mathbf{D}, d\mathbf{S}), \quad . \quad . \quad . \quad . \quad (19\cdot2)$$

where Q is the algebraic sum of the quantities introduced—or created or destroyed—within the boundary LMN, **D** is the displacement of the fluid at points on the boundary LMN, and the integration is extended over LMN. It is hardly necessary to point out that all this is in exact conformity with the results of experiments of the Faraday type. If a pressure difference be maintained between two points in the same cavity, *abc* for instance, a flow of the fluid from one point to the other will be maintained so long as the pressure difference is maintained ; whereas if such a pressure difference be established in a region simulating a dielectric the flow or displacement of the fluid will only continue (assuming the two points are not so close together that they are actually in the same cell) until the stresses evoked in the septa reach a certain value. There will then be no further

displacement of the fluid even though the pressure difference does not become zero.

We shall now leave our mechanical picture ; but will adopt the suggestion which is contained in it. This is the famous **displacement hypothesis** of Clerk Maxwell. We may state it in words as follows : **The generation of electric charges in any region is always associated with a displacement of electricity in such a way that the total (algebraic) quantity which passes outwards through any closed surface is equal to the algebraic sum of the charges established within it.** Its mathematical expression is contained in **(19·2)**, where Q means the algebraic sum of the charges within the closed surface over which the integration extends. We shall call the vector **D** the **electric displacement.** It is easy to see what it means. The scalar product

$$(\mathbf{D} \ \mathbf{dS}) = \mathbf{D} \cos \theta \ \mathbf{dS},$$

where θ is the angle between the direction of **D** and that of **dS**, represents the quantity of electricity which has passed through **dS** from one side to the other in the sense of the vector arrow of **dS**. In other words, **D means the quantity of electricity, reckoned per unit area, which is displaced in the direction of D through a small area dS, the surface of which is perpendicular to D.**

Let us suppose the surface in **(19·2)** to enclose a region occupied by a dielectric, or dielectrics, and that the charge density at any point is represented by ρ. Then the left-hand side of **(19·2)** may be written

$$\iiint \rho \ dx \ dy \ dz,$$

the integration extending over the enclosed region, while the right-hand side becomes

$$\iiint \mathbf{di}^{\cdot\prime} \mathbf{D} \ dx \ dy \ dz,$$

by the theorem of Gauss, **(3·01)**. We have consequently

$$\iiint \rho \ dx \ dy \ dz = \iiint \mathrm{div} \ \mathbf{D} \ dx \ dy \ dz.$$

If this has to hold for any volume, small or large, we must conclude that

$$\mathrm{div} \ \mathbf{D} = \rho. \quad . \quad . \quad . \quad . \quad . \quad \textbf{(19·21)}$$

From **(18·72)** and **(19·08)** we have further

$$\mathbf{div}\ \boldsymbol{\mathcal{E}} = \frac{4\pi A}{K}\rho, \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \textbf{(19·211)}$$

and we therefore infer that

$$\mathbf{D} = \frac{K}{4\pi A}\boldsymbol{\mathcal{E}}, \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \textbf{(19·22)}$$

since **D** must vanish with **ℰ**. One of the assumptions underlying **(19·211)** is that of the isotropic character of the dielectric, namely that the absolute value of the force between two point charges is independent of the orientation of the line joining them. How should the formula **(19·22)** be modified to meet the case of an anistropic dielectric ? The simplest generalization of it is the following :

$$\left.\begin{array}{l} 4\pi A\ D_x = K_{xx}\mathcal{E}_x + K_{xy}\mathcal{E}_y + K_{xz}\mathcal{E}_z, \\ 4\pi A\ D_y = K_{yx}\mathcal{E}_x + K_{yy}\mathcal{E}_y + K_{yz}\mathcal{E}_z, \\ 4\pi A\ D_z = K_{zx}\mathcal{E}_x + K_{zy}\mathcal{E}_y + K_{zz}\mathcal{E}_z, \end{array}\right\} \quad \cdot \quad \cdot \quad \textbf{(19·23)}$$

where the constants K_{xx}, K_{xy}, etc., are the components of a tensor of the second rank. We shall prove that **(19·22)** is a special case of **(19·23)**,[1] and we shall see later (**§ 25·4**) that these latter equations furnish a satisfactory foundation for the theory of the propagation of light (electromagnetic waves) in crystalline media.

We have already learned that the equations of transformation of the components of a tensor of the second rank are identical with those of the products of the components (taken two at a time) of two vectors (see **(2·22)** and **§ 2·3**). Thus, if **A** and **B** are any two vectors referred to a system of rectangular co-ordinates X, Y, Z, and if **A′** and **B′** are the same vectors referred to the system X', Y', Z', with the same origin as X, Y, Z ; then

$$A'_x B'_y = (\alpha_{11}A_x + \alpha_{12}A_y + \alpha_{13}A_z)(\alpha_{21}B_x + \alpha_{22}B_y + \alpha_{23}B_z)$$

or $\ A'_x B'_y = \alpha_{11}\alpha_{21}A_x B_x + \alpha_{11}\alpha_{22}A_x B_y + \alpha_{11}\alpha_{23}A_x B_z$

$$+\ \alpha_{12}\alpha_{21}A_y B_x + \alpha_{12}\alpha_{22}A_y B_y + \alpha_{12}\alpha_{23}A_y B_z$$

$$+\ \alpha_{13}\alpha_{21}A_z B_x + \alpha_{13}\alpha_{22}A_z B_y + \alpha_{13}\alpha_{23}A_z B_z.$$

Therefore

$$K'_{xy} = \alpha_{11}\alpha_{21}K_{xx} + \alpha_{11}\alpha_{22}K_{xy} + \alpha_{11}\alpha_{23}K_{xz}$$

$$+\ \alpha_{12}\alpha_{21}K_{yx} + \alpha_{12}\alpha_{22}K_{yy} + \alpha_{12}\alpha_{23}K_{yz}$$

$$+\ \alpha_{13}\alpha_{21}K_{zx} + \alpha_{13}\alpha_{22}K_{zy} + \alpha_{13}\alpha_{23}K_{zz}.$$

---

[1] That is to say, that the scalar quantity K is a special case of the tensor K_{xx}, K_{xy}, etc.

If now the components with unlike subscripts, K_{xy}, K_{xz}, etc., are all zero, and if

$$K_{xx} = K_{yy} = K_{zz} = K,$$

we must have

$$K'_{xy} = (\alpha_{11}\alpha_{21} + \alpha_{12}\alpha_{22} + \alpha_{13}\alpha_{23})K$$

or $$K'_{xy} = 0,$$

by **(2·24).**

Similarly, we may show that

$$
\begin{aligned}
K'_{xx} = {} & \alpha_{11}\alpha_{11}K_{xx} + \alpha_{11}\alpha_{12}K_{xy} + \alpha_{11}\alpha_{13}K_{xz} \\
& + \alpha_{12}\alpha_{11}K_{yx} + \alpha_{12}\alpha_{12}K_{yy} + \alpha_{12}\alpha_{13}K_{yz} \\
& + \alpha_{13}\alpha_{11}K_{zx} + \alpha_{13}\alpha_{12}K_{zy} + \alpha_{13}\alpha_{13}K_{zz}.
\end{aligned}
$$

Hence in the case we are assuming, namely where

$$K_{xx} = K_{yy} = K_{zz} = \mathrm{K},$$

and $$K_{xy} = K_{xz} = K_{yx}, \text{ etc. } = 0,$$

$$K'_{xx} = (\alpha_{11}{}^2 + \alpha_{12}{}^2 + \alpha_{13}{}^2)K$$

or $$K'_{xx} = K, \text{ by } \textbf{(2·23).}$$

Similarly, $$K'_{yy} = K'_{zz} = K.$$

Therefore the scalar dielectric constant may be regarded as a special case of a tensor dielectric constant.[1]

§ 19·3. The Displacement Ellipsoid

Let $\mathbf{D} \equiv (D_x,\ D_y,\ D_z)$ be the electric displacement at the origin of rectangular co-ordinates and $\boldsymbol{\mathcal{E}} \equiv (\mathcal{E}_x,\ \mathcal{E}_y,\ \mathcal{E}_z)$ the associated electric field intensity there. We shall represent the dependence of $\mathbf{D}$ on the value and direction of $\boldsymbol{\mathcal{E}}$ in another way. Imagine a straight line $\mathbf{r}$ drawn from the origin in the direction of $\boldsymbol{\mathcal{E}}$, so that

$$\frac{\mathcal{E}_x}{\mathcal{E}} = \frac{x}{r},\ \frac{\mathcal{E}_y}{\mathcal{E}} = \frac{y}{r},\ \frac{\mathcal{E}_z}{\mathcal{E}} = \frac{z}{r},\quad \cdot\quad\cdot\quad\cdot\quad \textbf{(19·3)}$$

where x, y and z are the components of $\mathbf{r}$. The first of the equations **(19·23)**, for example, may now be written

$$4\pi A D_x = \{K_{xx}x + K_{xy}y + K_{xz}z\}\frac{\mathcal{E}}{r},$$

---

[1] In the same way the pressure, p, in a liquid or gaseous medium may be regarded as a special case of a tensor p_{xx}, p_{xy}, etc.

or

$$\frac{4\pi A r}{\varepsilon}D_x = K_{xx}x + K_{xy}y + K_{xz}z$$

Similarly,

$$\frac{4\pi A r}{\varepsilon}D_y = K_{yx}x + K_{yy}y + K_{yz}z \qquad \cdot \quad \cdot \quad \cdot \quad \textbf{(19·31)}$$

$$\frac{4\pi A r}{\varepsilon}D_z = K_{zx}x + K_{zy}y + K_{zz}z$$

Now multiply these three equations respectively by x, y and z and add. We obtain

$$\frac{4\pi A r}{\varepsilon}(\mathbf{Dr}) = K_{xx}x^2 + K_{xy}xy + K_{xz}xz$$
$$+ K_{yx}yx + K_{yy}y^2 + K_{yz}yz$$
$$+ K_{zx}zx + K_{zy}zy + K_{zz}z^2. \quad \cdot \quad \cdot \quad \textbf{(19·32)}$$

The scalar product, $(\mathbf{Dr})$, is equal to $\mathbf{D}\cos\theta\, r$, where θ is the angle between the directions of $\mathbf{D}$ and $\mathbf{r}$, or between the directions of $\mathbf{D}$ and $\boldsymbol{\varepsilon}$. Therefore

$$(\mathbf{Dr}) = rD_n,$$

where D_n is the component of $\mathbf{D}$ in the direction of $\boldsymbol{\varepsilon}$. Substituting this in the left-hand side of (19·32) we get

$$4\pi A \frac{D_n}{\varepsilon} r^2 = K_{xx}x^2 + K_{xy}xy + K_{xz}xz$$
$$+ K_{yx}yx + K_{yy}y^2 + K'_{yz}yz$$
$$+ K_{zx}zx + K_{zy}zy + K_{zz}z^2. \quad \cdot \quad \textbf{(19·33)}$$

Let us suppose the vector $\mathbf{r}$ to have such a length, in all directions, that the product

$$r^2 D_n/\varepsilon$$

is constant, i.e. has the same value for all directions ; then (19·33) is the equation of a surface of the second degree. We shall call it the **displacement ellipsoid.** It obviously is an ellipsoid, since the radius vector $\mathbf{r}$ is positive and finite in all directions. This ellipsoid has the property that the ratio, D_n/ε, of the component of the displacement in the direction of $\boldsymbol{\varepsilon}$ to the value of the electric intensity, $\boldsymbol{\varepsilon}$, evoking it, is inversely proportional to the square of the radius vector of the ellipsoid in that direction.

By turning the co-ordinate system about the origin we can find new directions for the axes, such that (19·33) becomes

$$M = K_{xx}x^2 + K_{yy}y^2 + K_{zz}z^2, \quad \cdot \quad \cdot \quad \textbf{(19·331)}$$

where M is the constant value of $4\pi A r^2 D_n/\mathcal{E}$. In this system of co-ordinates all the components of the dielectric tensor except K_{xx}, K_{yy} and K_{zz} vanish. We may call these three non-vanishing components the **principal dielectric constants** or the **principal values of the dielectric constant.** The directions of the co-ordinate axes which reduce the equation of the displacement ellipsoid to the simplified form **(19·331)** we shall call the **principal axes of the displacement.** Referred to these axes equations **(19·23)** become :

$$4\pi A D_x = K_{xx}\mathcal{E}_x,$$

$$4\pi A D_y = K_{yy}\mathcal{E}_y, \quad . \quad . \quad . \quad . \quad . \quad \textbf{(19·34)}$$

$$4\pi A D_z = K_{zz}\mathcal{E}_z.$$

We see from **(19·23)** that $\mathbf{D}$ and $\mathcal{E}$ do not in general coincide in direction. For example, when the electric field intensity $\mathcal{E}$ is in the X direction, i.e. when $\mathcal{E}_y$ and $\mathcal{E}_z$ are both zero, we notice that D_y and D_z do not necessarily vanish. But, as equations **(19·34)** indicate, $\mathbf{D}$ and $\mathcal{E}$ coincide in direction when $\mathcal{E}$ is directed along a principal axis.

§ 19·4.   Polarization in Dielectric Media

It is convenient to distinguish, in the displacement $\mathbf{D}$, between the part of it due to, or associated with, the material medium, and the part of it due to the hypothetical aether. Thus we write

$$\mathbf{D} = \mathbf{D}_0 + \mathbf{P}, \quad . \quad . \quad . \quad . \quad . \quad \textbf{(19·4)}$$

where $\mathbf{D}$ is the displacement and $\mathbf{D}_0$ is the displacement which would exist at the same place with the same absolute value and direction of the field intensity $\mathcal{E}$, if the material medium were removed. With this meaning of $\mathbf{D}_0$, equation **(19·4)** defines the vector $\mathbf{P}$, which we shall call the **polarization** in the material medium. Confining our attention to isotropic media, we may write **(19·4)** in the form

$$\frac{K}{4\pi A}\mathcal{E} = \frac{K_0}{4\pi A}\mathcal{E} + \mathbf{P}, \quad . \quad . \quad . \quad \textbf{(19·41)}$$

where K_0 is the dielectric constant of empty space (aether). Therefore

$$\mathbf{P} = \frac{K - K_0}{4\pi A}\mathcal{E}. \quad . \quad . \quad . \quad \textbf{(19·411)}$$

When we adopt electrostatic units this becomes

$$P = \frac{K-1}{4\pi A}\mathcal{E}, \quad . \quad . \quad . \quad . \quad (19\text{·}412)$$

and when we adopt the ordinary or old-fashioned electrostatic units, for which the value unity is assigned to A,

$$P = \frac{K-1}{4\pi}\mathcal{E}. \quad . \quad . \quad . \quad (19\text{·}413)$$

§ 19·5.   Energy in Dielectric Media

A little consideration shows that an electrostatic field has energy of the potential kind [1] (§ 5·1). Consider two charged particles having, let us suppose, numerically equal charges of opposite sign. If left to the forces due to the associated field they will approach one another and acquire kinetic energy. As they approach more and more closely to one another the field will gradually vanish. By hypothesis the kinetic energy is gained at the expense of another type of energy while the field is disappearing. This is the (potential) energy of the field.

The mechanical picture of § 19·2 suggests that we should regard this energy as seated or localized in the dielectric medium. This way of regarding the energy is associated with the conception of the dielectric as being

Fig. 19·5

in a state of strain, of which the displacement is a measure. If this conception be sound a definite amount of work must be done *in each volume element* in producing in it an electric displacement or strain, and this work we must regard as a measure of the *energy in the element*. Let $ABCD$ (Fig. 19·5) be a cylindrical volume element in the dielectric, the cross-sectional area of which, AB, is equal to dS ; and let the fluid density in it be unity. Now suppose, in consequence of a field of force, a displacement of the fluid parallel to the axis of the cylinder, in the
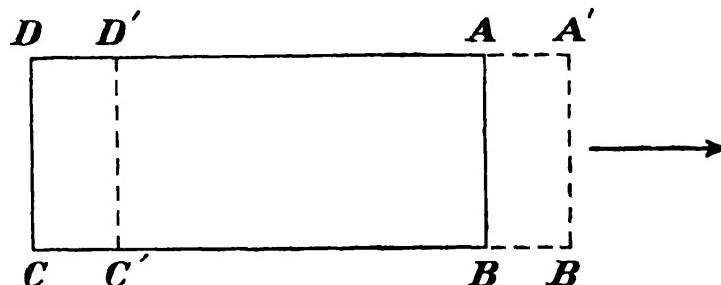
[1] The term ' potential ' suggests that the ' real ' energy is the kinetic variety, and no doubt the term was originally used to describe conditions capable of producing energy (in the sense of kinetic energy or *vis viva*), i.e. conditions in which energy was *potentially* present. This is not the modern usage of the term. Both *potential* and *kinetic* energy are on the same footing : one as ' real ' as the other.

sense indicated by the arrow, to be brought about. The spacial displacement is AA' ($= x$, shall we say) and the quantity of fluid displaced through AB or dS is equal to

$$(AA') \times dS = x dS.$$

Therefore the displacement per unit area is simply x. This is our vector $\mathbf{D}$. If $\mathcal{E}_x$ be the component (in the direction of $\mathbf{D}$, or x) of the force exerted on the unit quantity of the fluid, so that it corresponds to the x component of the electric field intensity, the total force, in the x direction, exerted on the fluid element $ABCD$ will be

$$\rho \times (\text{volume of cylinder } ABCD) \times \mathcal{E}_x,$$
$$= (\text{volume of cylinder } ABCD) \times \mathcal{E}_x,$$

since the density, ρ, is unity. The work done in effecting the displacement, x, is

$$v \int_0^x \mathcal{E}_x dx, \quad . \quad . \quad . \quad . \quad . \quad \textbf{(19·5)}$$

where v is the volume of the element. But $D_n/\mathcal{E}$ is a constant for a given direction n (§ 19·3), or, what amounts to the same thing, $\mathcal{E}_n/\mathbf{D}$ is a constant. In the present case x is the displacement, and consequently

$$\mathcal{E}_x = cx,$$

where c is the constant in question. The expression **(19·5)** therefore becomes

$$v \int_0^x cx dx$$

or

$$v \tfrac{1}{2} c x^2.$$

Hence the work done per unit volume is $\tfrac{1}{2} c x^2$. Now replace cx by $\mathcal{E}_x$ and x by $\mathbf{D}$ and the work done per unit volume, or the energy per unit volume, becomes

$$\tfrac{1}{2} \mathcal{E}_x \mathbf{D}.$$

This is identical with

$$\tfrac{1}{2} (\mathcal{E} \mathbf{D}), \quad . \quad . \quad . \quad . \quad . \quad . \quad \textbf{(19·51)}$$

since $\mathcal{E}_x$ is the component of the electric intensity in the direction of $\mathbf{D}$.

It will be noted that in deriving **(19·51)** we have not assumed the medium to be isotropic. In the special case of an isotropic medium $\mathbf{D} = K\mathcal{E}/4\pi A$, and **(19·51)** therefore becomes

$$E = \frac{K\mathcal{E}^2}{8\pi A}, \quad . \quad . \quad . \quad . \quad . \quad \textbf{(19·52)}$$

E being the energy per unit volume. When we adopt the ordinary type of units, for which $A = 1$, this becomes

$$E = \frac{K\mathcal{E}^2}{8\pi}, \quad . \quad . \quad . \quad . \quad \textbf{(19·521)}$$

and when we adopt Lorentz-Heaviside units, for which $\mathbf{A} = 1/4\pi$,

$$E = \frac{K\mathcal{E}^2}{2}. \quad . \quad . \quad . \quad . \quad \textbf{(19·522)}$$

§ 19·6. Energy as a Function of Charges and Potentials

The expressions **(19·51)** and **(19·52)** have been derived from a special picture of the electrostatic field which led us to regard the energy as seated or localized in the dielectric medium. We shall now endeavour to express the energy in an electrostatic field in terms of the charges and potentials at various points in the field. Imagine a number of elements of volume—we may take three of them as an illustration—in a dielectric medium (isotropic or anisotropic). For convenience they may be numbered 1, 2 and 3. Let us further imagine charges Q_1, Q_2 and Q_3 attached to the three volume elements respectively. The energy of the resulting electrostatic field will be equal to the work done in assembling these charges. If V_1, V_2 and V_3 be the respective potentials of the volume elements—we shall understand by potential in this connexion the amount of work, reckoned per unit charge, required to bring up an infinitesimal charge from an infinitely remote point and distribute it uniformly over the element—we shall have

$$\begin{aligned}
V_1 &= a_{11}Q_1 + a_{12}Q_2 + a_{13}Q_3, \\
V_2 &= a_{21}Q_1 + a_{22}Q_2 + a_{23}Q_3, \quad . \quad . \quad . \quad \textbf{(19·6)} \\
V_3 &= a_{31}Q_1 + a_{32}Q_2 + a_{33}Q_3,
\end{aligned}$$

where a_{11}, a_{12}, etc., are constants and $a_{xy} = a_{yx}$. We may justify these formulae in the following way : We assume the dielectric to be continuous in the sense of § 9·5, and we may therefore regard the electric density within each volume element

as uniform. The methods of **§ 19** show that, if we communicate some charge, Q_1, to the volume element 1, the resulting potential of the element is proportional to Q_1. Therefore in such a case

$$V_1 = a_{11}Q_1,$$

where a_{11} is a constant depending on the shape, position and (possibly) the orientation of the element. If, instead of this, we bring up some charge, Q_2, to volume element 2, the potential of 1 will become

$$V_1 = a_{12}Q_2,$$

where again a_{12} is a constant depending on the distance between 1 and 2 and, in the case of an anisotropic medium, on the orientation of the line joining the two elements. The two contributions together make V_1 equal to

$$V_1 = a_{11}Q_1 + a_{12}Q_2.$$

Similarly, if at the same time a charge Q_3 be given to element 3 the potential at 1 will be that expressed by the first of the equations **(19·6)**. The constants a_{xy} and a_{yx} are identical since each depends on the length and orientation of the same line, namely that joining the points x and y. It is now easy to calculate the work done in assembling the charges Q_1, Q_2 and Q_3, or in other words, the energy of the electrostatic field. It will be expressed by

$$E = \int_0^{Q_1} V_1 dQ_1 + \int_0^{Q_2} V_2 dQ_2 + \int_0^{Q_3} V_3 dQ_3.$$

Substituting in this equation the expressions **(19·6)** for V_1, V_2 and V_3 we get

$$E = \int_0^{Q_1} a_{11}Q_1 dQ_1 + \int_0^{Q_1} a_{12}Q_2 dQ_1 + \int_0^{Q_1} a_{13}Q_3 dQ_1$$

$$+ \int_0^{Q_2} a_{21}Q_1 dQ_2 + \int_0^{Q_2} a_{22}Q_2 dQ_2 + \int_0^{Q_2} a_{23}Q_3 dQ_2$$

$$+ \int_0^{Q_3} a_{31}Q_1 dQ_3 + \int_0^{Q_3} a_{32}Q_2 dQ_3 + \int_0^{Q_3} a_{33}Q_3 dQ_3.$$

We take the integrals in which different numerical subscripts appear, in pairs, thus

$$\int_{0}^{Q_1} a_{12}Q_2\,dQ_1 + \int_{0}^{Q_2} a_{21}Q_1\,dQ_2$$

$$= \int_{0}^{Q_1Q_2} a_{12}\,d(Q_1Q_2) = a_{12}Q_1Q_2,$$

$$= \int_{0}^{Q_1Q_2} a_{21}\,d(Q_2Q_1) = a_{21}Q_2Q_1,$$

since a_{12} and a_{21} are equal. Or, finally, we may write for the sum of this pair of integrals :

$$\tfrac{1}{2}a_{12}Q_1Q_2 + \tfrac{1}{2}a_{21}Q_2Q_1.$$

We see therefore that

$$\begin{aligned}
E = &\tfrac{1}{2}a_{11}Q_1{}^2 + \tfrac{1}{2}a_{12}Q_1Q_2 + \tfrac{1}{2}a_{13}Q_1Q_3 \\
&+ \tfrac{1}{2}a_{21}Q_2Q_1 + \tfrac{1}{2}a_{22}Q_2{}^2 + \tfrac{1}{2}a_{23}Q_2Q_3 \\
&+ \tfrac{1}{2}a_{31}Q_3Q_1 + \tfrac{1}{2}a_{32}Q_3Q_2 + \tfrac{1}{2}a_{33}Q_3{}^2,
\end{aligned}$$

and this, as reference to (19·6) shows, is equivalent to

$$E = \tfrac{1}{2}Q_1V_1 + \tfrac{1}{2}Q_2V_2 + \tfrac{1}{2}Q_3V_3.$$

The statement may obviously be generalized to apply to any number, n, of volume elements, so that

$$E = \tfrac{1}{2}\sum_{1}^{n}Q_sV_s. \quad . \quad . \quad . \quad . \quad . \quad \textbf{(19·61)}$$

Finally, if we regard the electricity as continuously distributed throughout the dielectric medium, the density at any point x, y, z being ρ, so that ρ is a function of x, y, z, we may express the associated energy in the form

$$E = \tfrac{1}{2}\iiint \rho V \, dx \, dy \, dz, \quad . \quad . \quad . \quad . \quad \textbf{(19·62)}$$

the integration being extended over all regions where changes exist.

Similarly, where we have charges assembled on surfaces (as they may be on conducting surfaces) we shall find for the associated energy the expression

$$E = \tfrac{1}{2}\iint \sigma V \, dS, \quad . \quad . \quad . \quad . \quad . \quad \textbf{(19·63)}$$

where σ is the surface density, and the integration is extended over all surfaces where charges appear.

§ 19·7. The Equilibrium of Conductors in an Electrostatic System—Quadrant Electrometer

In a mechanical system, in which there is a potential energy function independent of the time (**§ 5·1**), equilibrium can only exist when the potential energy is a maximum or a minimum. This is an immediate consequence of the principle of virtual displacements (**§ 8**); for

$$-\delta V = \underset{s}{\Sigma}(F_{sx}\delta x_s + F_{sy}\delta y_s + F_{sz}\delta z_s) = 0, \quad . \quad . \quad \textbf{(19·7)}$$

where V here means the **potential energy** of the system, and must not be confused with the rather similar electrostatic potential. It is easy to see that the *maximum* value of the potential energy is associated with *unstable* equilibrium. In fact in any departure from equilibrium the forces of the system do work *at the expense of the potential energy*, and so tend to make it as small as possible.

Charged conductors in an electrostatic field can only be kept in equilibrium by means of balancing forces of non-electric origin. When such balancing forces are associated with a mechanical potential energy of the kind described above, as for example in the quadrant electrometer where the deflexion of the needle is resisted by the twisting of the suspension, the whole system—electrical plus mechanical—*when it is in or near equilibrium*, comes within the scope of the above theorem; for the whole energy of the system then depends on its configuration : it is potential energy in the sense of **§ 5·1**. Stable equilibrium is therefore associated with a minimum value of the total energy, electrical plus mechanical.

We have been supposing the system to be isolated; i.e. that no energy has been transferred to it from external systems, or removed from it; and in the illustration of the electrometer we have tacitly assumed its electrostatic energy to remain unmodified by electrical connexions with external systems, i.e. we have assumed its conducting parts to be insulated. In the practical use of the instrument for measuring potential differences, the potentials of the needle and quadrants are, however, invariably determined by external connexions, and it constitutes a system which is not isolated. We shall therefore study the energy changes in a system of charged conductors during a *slow* motion from one configuration to another, *their potentials being main-*

tained constant. The restriction to slow motion enables us to regard the electrical field *as an electrostatic field.* Suppose the motion to be in the sense in which the conductors are impelled by the electrical forces ; so that the resisting mechanical forces are barely adequate for equilibrium. It is clear that, if the system were isolated the electrostatic energy would diminish, with a consequent change of the potentials ; since the charges would not be altered **(19·61).** If the potentials are to be kept constant it is therefore necessary that charges should be communicated to the conductors. Consider the case where the initial charges of the conductors are :

$$Q_1, \ Q_2, \ \ldots \ Q_n,$$

and their potentials :

$$V_1, \ V_2, \ \ldots \ V_n \text{ respectively.}$$

The final charges will be

$$Q_1 + q_1, \ Q_2 + q_2, \ \ldots \ Q_n + q_n,$$

the corresponding potentials being unchanged.

The initial electrostatic energy is

$$\tfrac{1}{2}\sum_{1}^{n} Q_s V_s$$

by **(19·61),** and the final energy is

$$\bullet \qquad \tfrac{1}{2}\sum_{1}^{n} (Q_s + q_s) V_s.$$

There is consequently an increment in the energy of the field equal to

$$\tfrac{1}{2}\sum_{1}^{n} q_s V_s. \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \textbf{(19·71)}$$

The energy supplied to the electrostatic field is clearly

$$\sum_{1}^{n} q_s V_s, \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \textbf{(19·72)}$$

since each V_s is constant and is equal to the work which must be done, at the expense of some external source of energy, to give the unit quantity of electricity to the conductor s.

We have not yet inquired whether the increments **(19·71)** and **(19·72)** are positive or negative, or whether either sign is possible. We are supposing the motion to be in the direction in which the *electrical* forces tend to impel the system. Therefore work is done at the expense of the electrical energy. Now,

as **(19·71)** and **(19·72)** indicate, the excess of the energy supplied to the field over the amount that actually remains in it is

$$\sum_1^n q_s V_s - \tfrac{1}{2}\sum_1^n q_s V_s$$

or $$\qquad \tfrac{1}{2}\sum_1^n q_s V_s. \quad . \quad . \quad . \quad . \quad \textbf{(19·721)}$$

Now this must clearly be equal to work done by the electrical forces, and consequently must be *positive*. Therefore **the work done by the electrical forces is equal to the increment in the field energy.** Or, if the work is done in producing mechanical potential energy, **the increment of the mechanical energy is equal to the increment of the field energy.** Both of these increments depend only on the initial and final configurations and are therefore independent of the way in which the system moves from one to the other ; so that the validity of the theorem is not confined to the special case of slow motion. It may be remarked, however, that the validity of the expression **(19·72)** is restricted to the case of slow motion.

The theorem just established can be applied to obtain the simple formula for the quadrant electrometer (see Starling, *Electricity and Magnetism*). Let the potentials of the needle and the two pairs of quadrants be respectively V, V_1 and V_2. We may suppose $V_2 > V_1$. The needle constitutes, with the two pairs of quadrants, a condenser; or, shall we say, two condensers : the portion of the needle overlapping the pair of quadrants with the potential V_1 being the one condenser, and that overlapping the other pair of quadrants being the other. A deflexion of the needle, θ, increases the capacity on one side at the expense of that on the other side, and the needle is so designed that, in the ideal case, $C\theta$ is the increase in capacity on one side, or the decrease on the other, due to the deflexion, θ : the coefficient, C, being a constant.

The deflexion of the needle is associated with a change of electrical energy *in the interior of the quadrants only* (vide § **18·1**). Let us consider first the energy associated with the condensing system consisting of the quadrants having the lower potential V_1 and the overlapping portion of the needle. By **(19·61)** this energy is

$$\tfrac{1}{2}Q_1 V_1 + \tfrac{1}{2}Q V,$$

Q being the charge on the overlapping portion of the needle. Now $Q_1 = - Q$ (§ **18·1**), so that we may write

$$\tfrac{1}{2}Q(V - V_1).$$

We know (§ 19) that Q is equal to the product of $V - V_1$ and the capacity of this part of the system. Therefore our expression for the energy becomes

$$\tfrac{1}{2} \times \text{Capacity} \times (V - V_1)^2,$$

and we have a similar expression for the energy of the other part of the condensing system.

Remembering that the increase in capacity on one side, or the diminution on the other, is $C\theta$, where θ is the angle through which the needle turns and C is a constant; we find for the net increase in the electrostatic energy due to a motion in the direction from 2 to 1 :

$$\tfrac{1}{2}C\theta(V - V_1)^2 - \tfrac{1}{2}C\theta(V - V_2)^2.$$

This must be equal to the corresponding increase in the mechanical energy, which is $\tfrac{1}{2}\tau\theta^2$, τ being the couple per unit angular twist of the suspension. Consequently

$$\tfrac{1}{2}\tau\theta^2 = \tfrac{1}{2}C\theta\{(V - V_1)^2 - (V - V_2)^2\},$$

or
$$\theta = k(V_2 - V_1)\left\{V - \frac{V_2 + V_1}{2}\right\}, \quad . \quad . \quad \textbf{(19·73)}$$

where k is a constant. When V is made equal to V_2, this becomes

$$\theta = k'(V_2 - V_1)^2, \quad . \quad . \quad . \quad \textbf{(19·731)}$$

where again k' is a constant.

These formulae are of course only approximate, since the assumptions we have made about the constitution of the electrometer do not hold exactly for any actual instrument. Indeed, in the form of quadrant electrometer devised by A. H. Compton, one of the quadrants can be raised or lowered, so that the symmetry of the system may be upset. This enables such an adjustment to be made as will cause the equilibrium to be very nearly neutral over the narrower range of potentials within which measurements are being carried out. Consequently small changes in $V_2 - V_1$ are associated with relatively enormous changes in θ. In other words, high sensitivity is secured.

Even if the instrument were so perfectly designed and adjusted that the formula (19·73) might be regarded as exact, it would not be easy to use it for absolute measurements of potential differences, on account of the difficulty of evaluating the constants C or k. In practice the instrument has to be calibrated by observing the deflexions produced by known potential differences.

§ 19·8. A THEOREM ON THE DISTRIBUTION OF CHARGE IN ELECTROSTATIC FIELDS

The expression **(19·61)** for the energy associated with a number of charged bodies may of course be written in the alternative forms :

$$\tfrac{1}{2}\sum_{1}^{n}C_s V_s{}^2 \quad \text{and} \quad \tfrac{1}{2}\sum_{1}^{n}\frac{Q_s{}^2}{C_s},$$

where C_s is the capacity of the body s ; since $Q_s = C_s V_s$ **(§ 19)**. Let us take the case of three conductors for example, and write down the formula for the difference of the energies associated with a given total charge, Q, in the two cases following :

(*a*) When the potentials of the conductors are V_1, V_2 and V_3, the respective charges being Q_1, Q_2 and Q_3, so that

$$Q = Q_1 + Q_2 + Q_3 ;$$

(*b*) when the charge Q is so distributed that the three potentials are equal to one another, i.e.

$$V_1 = V_2 = V_3 = V.$$

The energy in the case (*a*) is

$$E(a) = \tfrac{1}{2}C_1 V_1{}^2 + \tfrac{1}{2}C_2 V_2{}^2 + \tfrac{1}{2}C_3 V_3{}^2,$$

and in the case (*b*)

$$E(b) = \tfrac{1}{2}QV,$$

or

$$E(b) = \tfrac{1}{2}\frac{Q^2}{(C_1 + C_2 + C_3)},$$

and therefore

$$E(b) = \tfrac{1}{2}\frac{(C_1 V_1 + C_2 V_2 + C_3 V_3)^2}{C_1 + C_2 + C_3}.$$

Consequently

$$E(a) - E(b) = \tfrac{1}{2}C_1 V_1{}^2 + \tfrac{1}{2}C_2 V_2{}^2 + \tfrac{1}{2}C_3 V_3{}^2$$
$$- \tfrac{1}{2}\frac{(C_1 V_1 + C_2 V_2 + C_3 V_3)^2}{(C_1 + C_2 + C_3)}$$

and therefore

$$E(a) - E(b)$$
$$= \frac{1}{2C}\{(C_1 C_2 + C_1 C_3)V_1{}^2 + (C_2 C_1 + C_2 C_3)V_2{}^2 + (C_3 C_1 + C_3 C_2)V_3{}^2\}$$

$$- \frac{1}{2C}\{2C_1 C_2 V_1 V_2 + 2C_1 C_3 V_1 V_3 + 2C_2 C_3 V_2 V_3\},$$

where $\qquad\qquad\qquad C = C_1 + C_2 + C_3.$

Or finally,

$$E(a) - E(b)$$

$$= \frac{1}{2C}\{C_1 C_2 (V_1 - V_2)^2 + C_1 C_3 (V_1 - V_3)^2 + C_2 C_3 (V_2 - V_3)^2\}.$$

$$(19 \cdot 8)$$

We can easily generalize this expression so that it will apply for any number, n, of conductors. It then becomes

$$E_{(a)} - E_{(b)} = \frac{1}{4C}\overset{n}{\underset{1}{\Sigma}}_r \overset{n}{\underset{1}{\Sigma}}_s C_r C_s (V_r - V_s)^2. \qquad . \quad (19 \cdot 81)$$

The reason for the 4 in the denominator is that the summation counts every term (like $C_1 C_2 (V_1 - V_2)^2$ in **(19·8)**) twice over. It also includes terms $C_s C_s (V_s - V_s)^2$; but each of these is equal to zero.

The difference **(19·81)** is necessarily *positive or zero*, and hence the minimum energy is associated with such a distribution of charges as makes the potentials all equal. Charges of course tend to distribute themselves so as to bring about equality of potentials; because field intensities only vanish when the potential is everywhere the same. Hence we conclude that **the charges in an electrostatic field are distributed so as to give the field energy the minimum value possible.**

§ 19·9. ENERGY DENSITY IN DIELECTRIC MEDIA

Expressions for the energy density in dielectric media have already been obtained in **§ 19·5.** They were however deduced from a certain mechanical picture of the electrostatic field, and it is desirable to show that they are in harmony with the formulae of **§ 19·6.** The general expression for the energy in an electrostatic field can be written **(19·62 and 19·63)** in the form :

$$E = \tfrac{1}{2}\int\int \sigma V dS + \tfrac{1}{2}\int\int\int \rho V dx\, dy\, dz. \quad . \quad . \quad (19 \cdot 9)$$

In the further procedure it is convenient to give attention to the region within a closed conducting surface, such as abc in Fig. 19·9. We have already seen **(§ 18·1)** that such a surface completely disconnects, as it were, the interior from the exterior, and we are at liberty therefore to deal with the state of affairs within abc without referring at all to the region outside. We may suppose one or more conductors, def, within the space enclosed by abc and the rest of it occupied by dielectric media.

Our object is to prove, if possible, that the integral E **(19·9)** can be expressed in the form :

$$E = \tfrac{1}{2} \int\int\int (\mathcal{E}\mathbf{D}) \; dx \, dy \, dz, \quad . \quad . \quad . \quad \textbf{(19·91)}$$

the integration being extended over the volume occupied by dielectric. This volume is limited by the surface abc and by the surfaces def of any conductors lying within abc. The surface elements, $\mathbf{dS}$, of the boundary will have (in accordance with our usual practice) their vector arrows directed *outwards* from the region of integration, as shown for the elements 1 and 2 in Fig. 19·9. The electric displacement, $\mathbf{D}$, in the neighbourhood of a conducting surface such as abc or def has a component, D_n in the direction of $\mathbf{dS}$ (or of its vector arrow) equal *numerically* to the surface density, σ, but with the *opposite* sign, i.e.
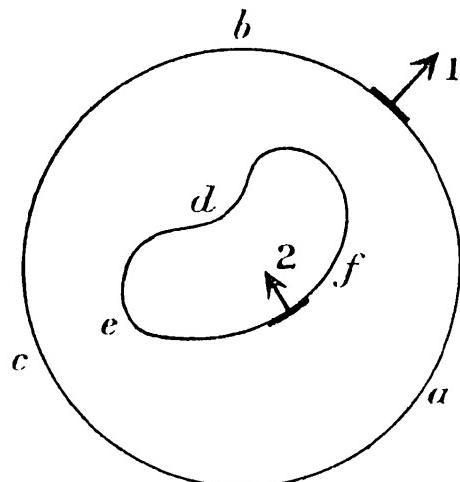
FIG. 19·9

$$D_n = -\,\sigma.$$

And as ρ is equal to $\mathbf{div}\,\mathbf{D}$, we may write **(19·9)** in the form :

$$E = -\,\tfrac{1}{2}\int\int V D_n \mathbf{dS} + \tfrac{1}{2}\int\int\int V\,\mathbf{div}\,\mathbf{D}\;dx\,dy\,dz,$$

or $\quad E = -\,\tfrac{1}{2}\int\int (V\mathbf{D},\;\mathbf{dS}) + \tfrac{1}{2}\int\int\int V\,\mathbf{div}\,\mathbf{D}\;dx\,dy\,dz\;;$

and consequently **(3·01)**

$$E = -\,\tfrac{1}{2}\int\int\int \mathbf{div}\,\{V\mathbf{D}\}\,dx\,dy\,dz + \tfrac{1}{2}\int\int\int V\,\mathbf{div}\,\mathbf{D}\;dx\,dy\,dz,$$

the integrations extending over the volume of the dielectric media.

It is now more convenient to employ the notation explained in § **2·4**, which enables us to write

$$\mathbf{div}\,\mathbf{A} \text{ in the form } (\nabla,\,\mathbf{A}),$$

where

$$\nabla \equiv (\nabla_x,\; \nabla_y,\; \nabla_z) \equiv (\partial/\partial x,\; \partial/\partial y,\; \partial/\partial z).$$

Thus our expression for the energy becomes

$$E = -\,\tfrac{1}{2}\int\int\int (\nabla,\; V\mathbf{D})\;dx\,dy\,dz + \tfrac{1}{2}\int\int\int V(\nabla\mathbf{D})\;dx\,dy\,dz.$$

But, as is easily verified,

$$(\nabla, V\mathbf{D}) = (\nabla V, \mathbf{D}) + V(\nabla \mathbf{D}).$$

Therefore

$$E = -\tfrac{1}{2}\iiint (\nabla V, \mathbf{D})\, dx\, dy\, dz.$$

Now $\nabla V = -\, \mathcal{E}$, consequently

$$E = \tfrac{1}{2}\iiint (\mathcal{E}\mathbf{D})\, dx\, dy\, dz.$$

This result applies to any dielectric medium, whether isotropic or not.

§ 20. MAXWELL STRESSES

Let F_x be the X component of the resultant force, reckoned per unit volume, exerted on a small portion of a dielectric medium. We have

$$F_x = \rho \mathcal{E}_x. \qquad \cdots \qquad (20)$$

The force we are dealing with is that due to the field only. This force will in general differ from zero, even when the medium is at rest ; though in this latter case the resultant of *all* the forces will necessarily be zero. We may write (20) in the form :

$$F_x = \mathcal{E}_x \operatorname{div} \mathbf{D}$$

by (19·21), or assuming that the dielectric is isotropic and homogeneous and that K is a constant,

$$F_x = \frac{K}{4\pi A}\mathcal{E}_x \operatorname{div} \mathcal{E}.$$

We have therefore

$$\frac{4\pi A}{K}F_x = \mathcal{E}_x\frac{\partial \mathcal{E}_x}{\partial x} + \mathcal{E}_x\frac{\partial \mathcal{E}_y}{\partial y} + \mathcal{E}_x\frac{\partial \mathcal{E}_z}{\partial z}$$

$$= \frac{\partial}{\partial x}\{\tfrac{1}{2}\mathcal{E}_x{}^2\} + \frac{\partial}{\partial y}\{\mathcal{E}_x\mathcal{E}_y\} + \frac{\partial}{\partial z}\{\mathcal{E}_x\mathcal{E}_z\}$$

$$-\, \mathcal{E}_y\frac{\partial \mathcal{E}_x}{\partial y} - \mathcal{E}_z\frac{\partial \mathcal{E}_x}{\partial z}. \qquad \cdots \qquad (20\cdot01)$$

Now

$$\mathcal{E}_x = -\frac{\partial V}{\partial x} \quad \text{and} \quad \mathcal{E}_y = -\frac{\partial V}{\partial y}.$$

Therefore

$$\frac{\partial \mathcal{E}_x}{\partial y} = \frac{\partial \mathcal{E}_y}{\partial x}.$$

The last two terms in **(20·01)** may therefore be written :

$$- \mathcal{E}_y \frac{\partial \mathcal{E}_y}{\partial x} - \mathcal{E}_z \frac{\partial \mathcal{E}_z}{\partial x},$$

or

$$\frac{\partial}{\partial x} \{ -\tfrac{1}{2}\mathcal{E}_y{}^2 - \tfrac{1}{2}\mathcal{E}_z{}^2 \} ;$$

so that **(20·01)** becomes

$$\frac{4\pi A}{K} F_x = \frac{\partial}{\partial x} \{ \tfrac{1}{2}\mathcal{E}_x{}^2 - \tfrac{1}{2}\mathcal{E}_y{}^2 - \tfrac{1}{2}\mathcal{E}_z{}^2 \} + \frac{\partial}{\partial y} \{\mathcal{E}_x \mathcal{E}_y\} + \frac{\partial}{\partial z} \{\mathcal{E}_x \mathcal{E}_z\}, \quad \textbf{(20·011)}$$

or

$$\frac{4\pi A}{K} F_x = \frac{\partial}{\partial x} \{\mathcal{E}_x{}^2 - \tfrac{1}{2}\mathcal{E}^2\} + \frac{\partial}{\partial y} \{\mathcal{E}_x \mathcal{E}_y\} + \frac{\partial}{\partial z} \{\mathcal{E}_x \mathcal{E}_z\}. \quad \textbf{(20·012)}$$

We may therefore express F_x, and the remaining components of the force per unit volume, in the form :

$$F_x = \frac{\partial t_{xx}}{\partial x} + \frac{\partial t_{xy}}{\partial y} + \frac{\partial t_{xz}}{\partial z},$$

$$F_y = \frac{\partial t_{yx}}{\partial x} + \frac{\partial t_{yy}}{\partial y} + \frac{\partial t_{yz}}{\partial z}, \quad . \quad . \quad . \quad \textbf{(20·02)}$$

$$F_z = \frac{\partial t_{zx}}{\partial x} + \frac{\partial t_{zy}}{\partial y} + \frac{\partial t_{zz}}{\partial z},$$

where the t's have the meanings :

$$t_{xx} = \frac{K}{4\pi A} \{\mathcal{E}_x{}^2 - \tfrac{1}{2}\mathcal{E}^2\}, \quad t_{xy} = \frac{K}{4\pi A} \mathcal{E}_x \mathcal{E}_y, \quad t_{xz} = \frac{K}{4\pi A} \mathcal{E}_x \mathcal{E}_z,$$

$$t_{yx} = \frac{K}{4\pi A} \mathcal{E}_y \mathcal{E}_x, \quad t_{yy} = \frac{K}{4\pi A} \{\mathcal{E}_y{}^2 - \tfrac{1}{2}\mathcal{E}^2\}, \quad t_{yz} = \frac{K}{4\pi A} \mathcal{E}_y \mathcal{E}_z, \quad \textbf{(20·03)}$$

$$t_{zx} = \frac{K}{4\pi A} \mathcal{E}_z \mathcal{E}_x, \quad t_{zy} = \frac{K}{4\pi A} \mathcal{E}_z \mathcal{E}_y, \quad t_{zz} = \frac{K}{4\pi A} \{\mathcal{E}_z{}^2 - \tfrac{1}{2}\mathcal{E}^2\}.$$

Reference to **§ 9·9** leads to the suggestion that the force **F** can be ascribed to a system of stresses in the dielectric, described by the tensor **t**. The components t_{xx}, t_{yy} and t_{zz} are like tensions along lines parallel to the X, Y and Z axes respectively ; while t_{xy}, t_{xz}, t_{yz}, etc., are analogous to shearing stresses **(§ 9·7)**. The stresses described by the tensor **(20·03)** are known as **Maxwell stresses**, after Clerk Maxwell, to whom this way of describing the mechanical forces in an electrostatic (or electromagnetic) field is due.

§ 20·1. Nature of the Maxwell Stresses

Let us suppose the co-ordinate axes turned so that the X axis has the same direction as the electric field intensity, $\mathcal{E}$, at some point (x, y, z). At this point then

$$\mathcal{E}_x = \mathcal{E},$$
$$\mathcal{E}_y = 0,$$
$$\mathcal{E}_z = 0 ;$$

and the stress tensor, $\mathbf{t}$, becomes

$$\frac{K}{8\pi A}\mathcal{E}^2, \qquad 0, \qquad 0,$$
$$0, \quad -\frac{K}{8\pi A}\mathcal{E}^2, \qquad 0,$$
$$0, \qquad 0, \quad -\frac{K}{8\pi A}\mathcal{E}^2,$$

as reference to (20·03) will show. It thus appears that the state of stress at any point may be described as a **tension** along the lines of force equal to $\frac{K}{8\pi A}\mathcal{E}^2$ and **tensions** along lines perpendicular to the lines of force equal to $-\frac{K}{8\pi A}\mathcal{E}^2$, i.e. **pressures** along lines perpendicular to the lines of force equal to $+\frac{K}{4\pi A}\mathcal{E}^2$.

Or (§ 19·5) we may describe the state of stress as **a tension along the lines of force equal to the energy per unit volume** at the point in question and **a pressure along lines perpendicular to the lines of force likewise equal to the energy per unit volume.**

Our notion of a state of stress has been derived (§ 9·7) from the study of the forces evoked by the deformation of an elastic medium. As an illustration, let us take the component, t_{xy}, of the stress tensor, (9·721), which is described by Fig. 9·73 and the accompanying text. It is really a measure of the resisting capacity of the medium to a certain kind of deformation, namely a shear. Now a fluid medium, in equilibrium, offers *no* resistance to this particular type of deformation (§ 10·1), and components of the stress tensor like t_{xy} are necessarily zero. But the t_{xy} of the Maxwell tensor (20·03) is in general different from zero, even in a fluid dielectric; since it is determined essentially by the product $\mathcal{E}_x\mathcal{E}_y$, the only physical characteristic of the medium which has any influence on it being the dielectric constant.

The explanation of the paradox is that, in using the term **Maxwell stresses,** we have tacitly widened the definition of **stress.** We have, in effect, agreed that the state of affairs described by a tensor, **t,** of the second rank, shall be called **stress** when the force per unit volume is given by equations **(20·02),** and we shall continue to use this broader definition of stress. But **the Maxwell stresses are not stresses at all in the narrow sense of § 9·7.** Indeed, it is quite clear that **the forces immediately due to an electrostatic field are exerted only on the charged portions of the material media, and mechanical deformation, with associated elastic stresses, can only arise as a secondary phenomenon ;** as, for example, when the tractive force between the plates of a condenser brings about a compression of the glass or mica separating them.

§ 20·2. THE ABSOLUTE ELECTROMETER

Imagine two conductors, a and b

(a)

(b)

FIG. 20·2

(Fig. 20·2 (A)), with plane parallel surfaces, cd and ef, the intervening space being occupied by an isotropic dielectric medium, g ; and suppose a uniform electrostatic field, $\mathcal{E}$, maintained in the dielectric, perpendicular to cd or ef, as illustrated in the figure. The investigations in the preceding section show that there will be a force, equal to $K\mathcal{E}^2/8\pi A$ on the unit area, tending to draw the two surfaces together. The field being uniform,

$$\mathcal{E} = (V_a - V_b)/d,$$
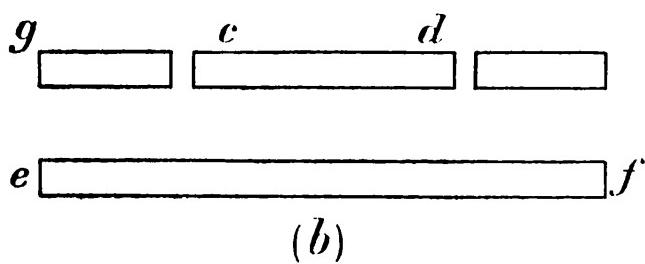
where V_a and V_b are the potentials of a and b respectively, and d is the distance between the two conducting surfaces. The tractive force over an area S of either surface is consequently

$$F = \frac{KS(V_a - V_b)^2}{8\pi A d^2}. \quad . \quad . \quad . \quad (20\cdot2)$$

We have in this formula the theoretical basis of the absolute electrometer (attracted disc electrometer) devised by Lord

Kelvin. The instrument, in its original form, consisted of a plate ef (Fig. 20·2 (B)) which could be raised or lowered by a micrometer screw, so that the distance, d, separating it from a parallel plate, cd, could be varied at will. This latter plate is supported by a system of springs, so that ordinarily its lower surface is slightly above that of a guard ring, g. The force necessary to maintain it in what we shall call its normal position, with its lower surface flush with that of the guard ring, can be determined by placing weights on it, while no potential difference exists between cd and the guard ring on the one hand, and ef on the other. This gives the force F of **(20·2)**. The weights having been removed, the potential difference to be measured is applied between cd and ef, the guard ring having of course the same potential as cd. The micrometer screw is then adjusted so that the lower surface of cd is again brought into the normal position. Finally, the potential difference is removed and the distance, d, separating the plates measured by the micrometer screw. The difference of potential can now be determined absolutely by formula **(20·2)** from the known values of F, S, d and K (for air). Another way of using the instrument is to apply some fixed potential, V, to cd and the guard ring. The potentials V_a and V_b, the difference of which is to be determined, are applied separately to the lower plate ef. The distances d_a and d_b corresponding to the differences $V_a - V$ and $V_b - V$ respectively do not require to be measured separately, but only their difference $d_a - d_b$. By **(20·2)** we have

$$V_a - V = \sqrt{\frac{8\pi A F}{KS}}\, d_a,$$

and

$$V_b - V = \sqrt{\frac{8\pi A F}{KS}}\, d_b.$$

Hence

$$V_a - V_b = \sqrt{\frac{8\pi A F}{KS}}\, (d_a - d_b). \quad . \quad . \quad \textbf{(20·21)}$$

This mode of using the electrometer (employing formula **20·21**) is preferable to the more direct one. If, for example, the planes of the disc, cd, and the guard ring, g, did not coincide, or if they were not exactly parallel to that of the other plate, ef, the formula **(20·2)** would cease to represent the potential difference $V_a - V_b$, and indeed the distance, d, would become vague. On the other hand, the consequent error in the formula **(20·21)** would be of a smaller order, while the difference $d_a - d_b$ remains quite definite and can be accurately measured.

CHAPTER III

ELECTROSTATIC FIELDS IN REGIONS CONTAINING SEVERAL DIELECTRIC MEDIA

§ 20·3. BOUNDARY CONDITIONS

WHILE we were still dealing with a single isotropic dielectric medium we learned that the integral, $\int (\mathcal{E}, d\mathbf{l})$, between two given points is independent of the path of integration, and consequently if it be taken round a closed loop, the result is zero. Therefore

$$\oint (\mathcal{E}, d\mathbf{l}) = 0 \quad . \quad . \quad . \quad . \quad (20\text{·}3)$$

(see **(18·231)** and also **§ 18·8)**. We shall assume that **(20·3)** holds for any closed path in any electrostatic field—even when there are several different dielectric media, isotropic or otherwise. The assumption will be found **(§ 22·4)** to be a special case of a more general law which we shall investigate later.
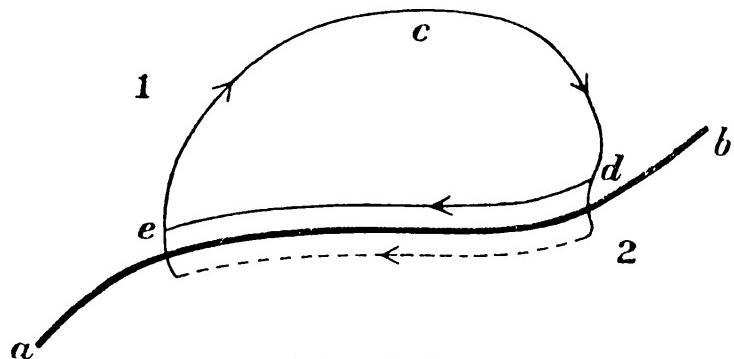
FIG. 20·3

Let ab be a portion of the boundary surface between two dielectric media 1 and 2 (Fig. 20·3). We may imagine it to cut the plane of the paper in the line ab. The integral $\oint(\mathcal{E}d\mathbf{l})$ round any closed loop, cde, is zero. Consider the two cases : (1) the whole of the loop is in medium 1, a portion of it, de, running along the boundary ; (2) the portion of the loop, de, which runs along the boundary, is in medium 2, and is practically coincident with the portion de in medium 1. We have consequently

$$\int_{ecd} (\mathcal{E}d\mathbf{l}) + \int_{de \text{ in } 1} (\mathcal{E}d\mathbf{l}) = 0,$$

64

and

$$\int_{ecd} (\mathcal{E}dl) + \int_{de\ in\ 2} (\mathcal{E}dl) = 0,$$

and therefore

$$\int_{de\ in\ 1} \mathcal{E}dl) = \int_{de\ in\ 2} (\mathcal{E}dl). \quad . \quad . \quad . \quad . \quad (20\cdot31)$$

We may write this result in the form :

$$\int_{de\ in\ 1} \mathcal{E}_p dl = \int_{de\ in\ 2} \mathcal{E}_p dl, \quad . \quad . \quad . \quad (20\cdot311)$$

where $\mathcal{E}_p$ in each case means the component of $\mathcal{E}$ parallel to dl. Since **(20·311)** must hold however short the path dl may be, we see that

$$\mathcal{E}_p \text{ (in medium 1)} = \mathcal{E}_p \text{ (in medium 2)}. \quad . \quad . \quad (20\cdot32)$$

In words : **the component of the electric field intensity at a boundary point, in a direction parallel to the boundary, has the same value on both sides of the boundary.**

We have seen **(§ 18·2)** that this is also the case when the boundary separates a dielectric medium from a conducting one ; in which case it happens that the component parallel to the boundary has the value zero on both sides.

We can obtain another boundary condition with the help of Maxwell's displacement hypothesis. In Fig. 20·31 let *ab* be a portion of a surface (perpendicular to the plane of the paper) separating two dielectric media. Let *cdef* be a small flat cylinder with its axis perpendicular to the surface *ab*. We shall suppose the portion of the surface *ab* in the figure to be so small that it may be regarded as plane. The cylinder is contained by the surfaces *cd* and *ef* which are plane and parallel to *ab*, and the curved surface, *cf—ed*,

FIG. 20·31

cutting *ab* in a closed loop, the area of which is equal to that of *cd* or *fe*. We may suppose *cf* or *de* to be so short that the area of the curved cylindrical surface is infinitesimal by comparison with that of either of the flat surfaces *cd* or *fe*. Let us now apply formula **(19·2)**, which we adopt as an axiom

5
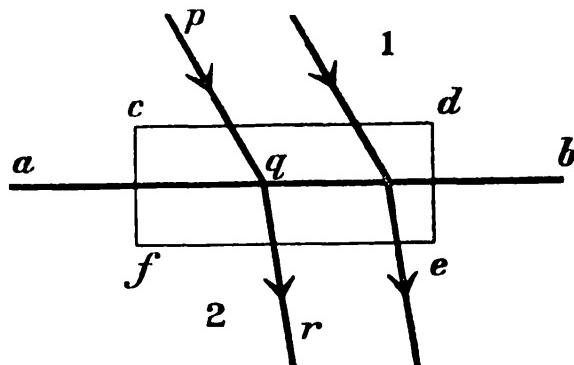
of universal validity, assuming that no changes are present near the boundary. We have

$$\int\int (\mathbf{D},\ \mathbf{dS}) = 0,$$

the integration extending over the whole surface of the cylinder *cdef*. Since, however, the curved portion of the cylindrical surface is negligibly small, this may be written

$$(D_n \times \text{area } cd)_1 + (D_n \times \text{area } fe)_2 = 0,$$

in which it will be remembered that n is directed *outwards* from the cylindrical enclosure, i.e. in the case of *cd*, *upwards* and, in the case of *fe*, *downwards*. When the electric displacement is across the boundary from 1 to 2, it is convenient to take the direction of n to be downwards for both *cd* and *fe*, in which case

$$(D_n \times \text{area } fe)_2 - (D_n \times \text{area } cd)_1 = 0,$$

or

$$(D_n)_1 = (D_n)_2. \quad . \quad . \quad . \quad . \quad \textbf{(20·34)}$$

We may express this in words as follows : **The component of the electric displacement in the direction of a normal at any point on a surface separating two dielectric media has the same value in the immediate neighbourhood of this point on both sides of the surface.**

Turning to the special case where the surface separates two different isotropic dielectrics (1 and 2 in Fig. 20·31), equation **(20·34)** becomes

$$\frac{K_1}{4\pi A}\mathcal{E}_{1n} = \frac{K_2}{4\pi A}\mathcal{E}_{2n},$$

or

$$K_1\mathcal{E}_{1n} = K_2\mathcal{E}_{2n}, \quad . \quad . \quad . \quad . \quad \textbf{(20·35)}$$

where $\mathcal{E}_{1n}$ and $\mathcal{E}_{2n}$ are the values of the component of the electric intensity at q in the media 1 and 2 respectively in the direction of the normal from 1 to 2. The equation may be written

$$K_1\mathcal{E}_1 \cos \theta_1 = K_2\mathcal{E}_2 \cos \theta_2, \quad . \quad . \quad \textbf{(20·351)}$$

where θ_1 and θ_2 may be described as the angles of incidence and refraction of the lines of force at the boundary surface.

A plane containing the normal at a boundary point, and parallel to $\mathcal{E}_1$, we may term a plane of incidence. Similarly, we may term a line of force in 1 passing through this boundary point (point of incidence) an incident line of force. The component of the incident field, at the point of incidence, is zero

in a direction normal to the plane of incidence ; and, in consequence of **(20·32)**, so also is the corresponding component of the field intensity at this point in medium 2. In other words, the plane containing the refracted line of force coincides with the plane of incidence. It follows that equation **(20·32)** may be written

$$\mathcal{E}_1 \sin \theta_1 = \mathcal{E}_2 \sin \theta_2. \quad . \quad . \quad . \quad (20\text{·}36)$$

On dividing **(20·36)** by **(20·351)** we get

$$\frac{\tan \theta_1}{K_1} = \frac{\tan \theta_2}{K_2}, \quad . \quad . \quad . \quad (20\text{·}37)$$

which expresses the law of refraction of the lines of force at a surface separating two isotropic dielectric media.

The validity of the formula **(20·35)** is not confined to the case where the boundary separates two dielectric media. It must obviously hold when one of the media, say medium 2, is a conductor ; in which case $\mathcal{E}_{2n}$ is zero. In such a case therefore K_2 must be infinite. A conductor therefore behaves (in an electrostatic field) like a dielectric which has an infinite dielectric constant.

§ 20·4. CONDUCTING SPHERE IN AN ELECTROSTATIC FIELD— ELECTRICAL IMAGES

We shall now study the problem of the electrostatic field due to a point charge, e (Fig. 20·4), with a conducting sphere in its neighbourhood. Let us suppose the region under investigation to be enclosed within an infinitely remote conducting surface, the potential of which we shall take to be zero, and that the conducting sphere has acquired the potential zero by having been temporarily in conducting connexion with the distant surface. The rest of the
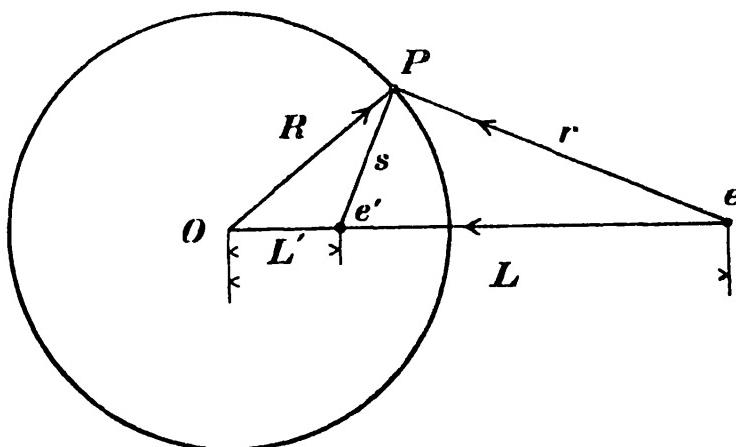
Fig. 20·4

region is occupied by an isotropic dielectric medium of which the dielectric constant is K. We are given then that the potential at all points in the conducting sphere, and in particular at all points on its boundary, is zero. We are interested in the

resulting electrostatic field outside the sphere and in the charges induced on its surface.

Lord Kelvin showed that this and similar problems could be solved by the following artifice : We imagine the sphere to be removed and its place filled with the dielectric medium the constant of which is K. We can establish the potential zero at all points on the spherical boundary by placing a suitable point charge, e', at a certain point within the sphere on the line Oe (Fig. 20·4). Let the distances of e and e' from O, the centre of the spherical region, be L and L' respectively and let r and s be their respective distances from any point, P, on the spherical surface. Then we require, in accordance with **(18·7)**, that

$$\frac{Ae}{Kr} + \frac{Ae'}{Ks} = 0$$

shall hold at all such points. Consequently

$$\frac{e'}{e} = -\frac{s}{r}.$$

The left-hand member of this equation is a constant and consequently s/r must be a constant. We can meet this requirement by choosing the position of e' so that it makes the triangles OPe and $Oe'P$ similar to one another, and consequently

$$\frac{s}{r} = \frac{R}{L} = \frac{L'}{R}.$$

Therefore

$$e' = -\frac{R}{L}e, \quad . \quad . \quad . \quad . \quad . \quad \textbf{(20·4)}$$

and $$R^2 = LL'. \quad . \quad . \quad . \quad . \quad . \quad \textbf{(20·41)}$$

So that a charge, e', if placed at a distance L' from the spherical centre, O, will, if the relations **(20·4)** and **(20·41)** hold, produce zero potential at all points on the spherical surface and the field *outside* the spherical region will be identical with that which was there before the replacement of the conducting sphere. The calculation of the field intensities outside the sphere has now become very easy, since the problem is reduced to one of point charges. Kelvin called the point charge, e', an **electrical image.**

Let us calculate the intensity at a point, P (Fig. 20·4), in the immediate neighbourhood of the spherical surface. We can

shorten our work by making use of the known fact that the intensity is normal to the surface. The intensity at P, due to e, is

$$\frac{Ae}{Kr^2}$$

and is in the direction eP. Imagine it replaced (in accordance with the triangle of vectors) by two intensities, one directed parallel to eO and the other along OP. The former we shall represent by $\mathcal{E}'$, while the latter is obviously equal to

$$\frac{Ae}{Kr^2} \times \frac{R}{r}.$$

The corresponding contributions of the image e' are an intensity parallel to eO, which we may call $\mathcal{E}''$, and one along the outward normal OP, equal to

$$\frac{Ae'}{Ks^2} \times \frac{R}{s}.$$

The contributions of the two images give for the intensity at P

$$\left. \begin{aligned} \frac{AeR}{Kr^3} + \frac{Ae'R}{Ks^3} & \text{[Direction of outward normal],} \\ \mathcal{E}' + \mathcal{E}'' \quad & \text{[Direction parallel to } eO]. \end{aligned} \right\} \quad \textbf{(20·42)}$$

The latter of these contributions we know already is equal to zero, and the resultant intensity reckoned along the outward normal is therefore given by the first of the expressions **(20·42)**. On substituting in it the appropriate expressions for s and e' we easily find

$$\mathcal{E} = \frac{Ae}{Kr^3R}(R^2 - L^2), \quad . \quad . \quad . \quad \textbf{(20·43)}$$

where $\mathcal{E}$ is the intensity, reckoned outwards, at P.

We may say of the result **(20·43)** that when it holds at all points on the conducting sphere *on the outside* the condition of zero potential is satisfied, or more generally that the external field calculated from e and the image e' placed in the position described above, constitutes a solution of the given problem ; but we are still in doubt as to whether some other mode of attack might not lead to a different result, which might nevertheless satisfy the given conditions, and in particular that of zero potential over the spherical surface. We shall prove in § 20·9 that there is only *one* solution of the problem, and that consequently it must be the one we have found.

It is easy to determine the charge density, σ, at any point, P, on the sphere. By (18·81) we have

$$\frac{4\pi A \sigma}{K} = \mathcal{E},$$

and consequently (20·43)

$$\sigma = -\frac{e}{4\pi r^3 R}(L^2 - R^2). \quad . \quad . \quad . \quad (20\text{·}45)$$

The force exerted on the sphere by the point charge, e, is evidently the same as that exerted by e on e'.

§ 20·5. Insulated Uncharged Spherical Conductor in the Field of a Point Charge

Here the conditions to be satisfied are : (*a*) constancy of potential over the spherical surface and (*b*) the algebraic sum of the number of lines of force leaving the spherical region has to be zero. We can meet these conditions by retaining the image e' exactly as in the last section, and introducing an extra image, e'', such that

$$e' + e'' = 0,$$

and consequently

$$e'' = -\frac{Re}{L}. \quad . \quad . \quad . \quad . \quad (20\text{·}5)$$

Clearly this extra image must be placed at the centre of the sphere in order that the potential at its surface may be constant. Since e' *alone* would reduce the potential to zero at all points on the spherical surface, it is clear that the potential is that due to e''. It must therefore be

$$\frac{Ae''}{KR} = \frac{Ae}{KL}, \quad . \quad . \quad . \quad . \quad (20\text{·}51)$$

by (20·5).

To get the intensity (in the outward direction) at points in the immediate neighbourhood of the surface of the sphere we have merely to add the expression

$$\frac{Ae''}{KR^2} \quad \text{or} \quad \frac{Ae}{KRL}$$

to (20·43). We thus obtain

$$\mathcal{E} = \frac{Ae}{KRL} + \frac{Ae}{Kr^3R}(R^2 - L^2), \quad . \quad . \quad (20\text{·}52)$$

and for the surface density at any point, P,

$$\sigma = \frac{e}{4\pi RL} + \frac{e}{4\pi r^3 R}(R^2 - L^2). \quad . \quad . \quad (20\cdot53)$$

§ 20·6. Charged Sphere and Point Charge

Again we consider an insulated spherical conductor in the neighbourhood of which is a point charge, e; but this time we shall suppose the sphere to have a charge Q. On replacing the sphere by dielectric, we have to secure that the potential shall be constant at all points on the spherical surface, and that the number of lines emerging through the spherical surface shall be that corresponding to the charge Q. The algebraic sum of the image charges must therefore be Q. It is easy to see that we satisfy these conditions by images e' and e'' equal to those described in the preceding sections, and having the same situations, and the additional image, equal to Q, placed at the spherical centre. We find, in consequence, for the constant potential over the spherical surface the value (20·51) plus that due to the point charge Q, or

$$V = \frac{AQ}{KR} + \frac{Ae}{KL}, \quad . \quad . \quad . \quad . \quad (20\cdot6)$$

and for the intensity we have the expression (20·52) plus that due to Q, or

$$\mathcal{E} = \frac{AQ}{KR^2} + \frac{Ae}{KRL} + \frac{Ae}{Kr^3 R}(R^2 - L^2). . \quad . \quad (20\cdot61)$$

Finally

$$\sigma = \frac{Q}{4\pi R^2} + \frac{e}{4\pi RL} + \frac{e}{4\pi r^3 R}(R^2 - L^2). . \quad . \quad (20\cdot62)$$

The mechanical force exerted by the charged sphere on the point charge e can easily be calculated. It is obviously equal to the resultant force exerted on e by the images e', e'' and Q. The derivation of an expression for this force may be left to the reader. If the point charge, e, be in the immediate neighbourhood of the spherical surface the force becomes *an attraction*, even when the algebraic sum of the charges on the sphere has the same sign as e. The force of attraction becomes in the limit when the distance, x, between e and the spherical surface approaches zero,

$$\frac{Ae^2}{4Kx^2}.$$

§ 20·7. Field of a Dipole

In Fig. 20·7, A and B represent two point charges each numerically equal to e, the former being negative and the latter positive. They are separated by a short distance which we shall represent by $2l$. Such a combination of charges is called an **electric doublet** or **dipole**. The product of the positive point charge and the distance separating the two point charges is called the **moment** of the dipole. If we denote it by **M**,
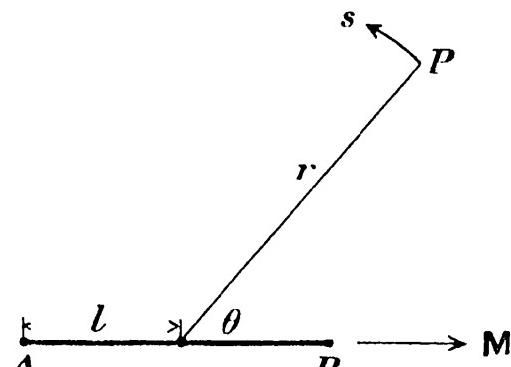
Fig. 20·7

$$\mathbf{M} = 2el. \quad . \quad . \quad . \quad . \quad \textbf{(20·7)}$$

We regard **M** as a vector, assigning to it the direction AB from the negative to the positive charge.

The potential, V, at a point, distant r from the centre of the dipole is

$$V = \frac{A}{K}\left(\frac{e}{r_1} - \frac{e}{r_2}\right),$$

where r_1 and r_2 are the distances from B and A respectively to P. This becomes in the limit, when $2l$ is very small,

$$V = \frac{A}{K}\frac{\mathbf{M}\cos\theta}{r^2}. \quad . \quad . \quad . \quad . \quad \textbf{(20·71)}$$

The component of the electric intensity at P, in the direction of the line r from the dipole to P, is

$$-\frac{\partial V}{\partial r} = \frac{2A\mathbf{M}\cos\theta}{Kr^3}. \quad . \quad . \quad . \quad \textbf{(20·72)}$$

Let s represent distances measured from P in a direction at right angles to the radial line r as shown in Fig. 20·7. The electric intensity in such a direction will be

$$-\frac{\partial V}{\partial s} = -\frac{\partial V}{r\partial \theta} = \frac{A\mathbf{M}\sin\theta}{Kr^3}. \quad . \quad . \quad \textbf{(20·721)}$$

§ 20·8. Conducting Sphere in a Uniform Field

We turn now to the case of an uncharged conducting sphere situated in a field which, before the introduction of the sphere,

was uniform, i.e. everywhere constant in magnitude and direction. Suppose the intensity to be $\mathcal{E}$. We obtain at once an approximation to the solution of this problem by adopting the results of § 20·5 and supposing

$$L \gg R.$$

We shall then have

$$\mathcal{E} = \frac{Ae}{KL^2}. \quad . \quad . \quad . \quad . \quad . \quad (20\cdot 8)$$

Let us replace e in the formulae of § 20·5 by means of (20·8), and allow L to approach the limit ∞. We find

$$e' = -\frac{KRL}{A}\mathcal{E},$$

$$e'' = +\frac{KRL}{A}\mathcal{E}.$$

The two images will approach one another infinitely closely, since $LL' = R^2$, and will constitute a dipole the moment of which is easily seen to be

$$\mathbf{M} = \frac{KR^3\mathcal{E}}{A}. \quad . \quad . \quad . \quad . \quad (20\cdot 81)$$

The direction of $\mathbf{M}$ is that of the field $\mathcal{E}$.

We may regard the field outside the sphere as a superposition of the field due to the dipole, the moment of which is expressed by (20·81), on the original field $\mathcal{E}$, or, alternatively, we may use the formulae of § 20·5, replacing e by $KL^2\mathcal{E}/A$ and allowing L to approach ∞.

We know that the electric intensity just outside the sphere has a direction normal to its surface. Formula (20·72)

FIG. 20·8

gives us, for the contribution of the dipole, the expression $\dfrac{2AM\cos\theta}{KR^3}$ in an outward direction, or, when we substitute $KR^3\mathcal{E}/A$ for $\mathbf{M}$ (20·81),

$$2\mathcal{E}\cos\theta \;;$$

(Fig. 20·8) while the contribution of the original field, $\mathcal{E}$, is

$$\mathcal{E}\cos\theta.$$
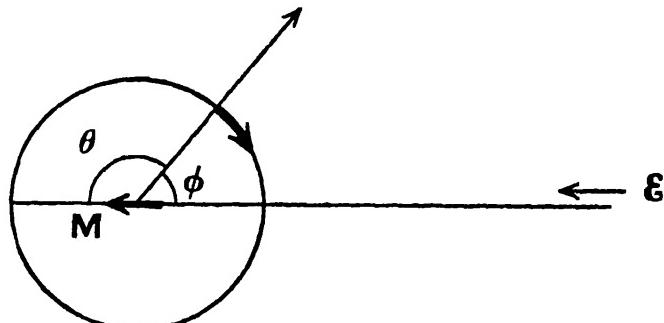
Therefore the electric intensity at the surface of the sphere, in an outward sense, is

$$3\mathcal{E}\cos\theta.\quad.\quad.\quad.\quad.\quad.\quad\textbf{(20·82)}$$

The surface density, σ, is given by multiplying this expression by $K/4\pi A$, so that we obtain

$$\sigma = \frac{3K\mathcal{E}\cos\theta}{4\pi A}.\quad.\quad.\quad.\quad.\quad\textbf{(20·83)}$$

§ 20·85.   Uniqueness of the Foregoing Solutions

The methods adopted in **§ 20·4** and in the following sections are such that they leave us in doubt as to whether the solution found is the only one. We can prove, however, that each of these solutions is unique. We shall confine our attention to the problem of **§ 20·4** ; but the same method will establish the uniqueness of the solutions of the other problems. We may regard the dielectric medium as bounded by a distant surface the potential over which is everywhere zero, and by the surface of the sphere over which the potential is likewise zero. We may also imagine a sphere of minute radius, ρ, with the point charge, e, as centre. The potential over this spherical surface is also given. It is in fact $Ae/K\rho$. If we give our attention to the region outside the spheres of radii R and ρ respectively, we may say that it is bounded by surfaces where the potential is given. Now the solution we have found enables us to assign a value, V, to the potential at any point in the region, and it gives, as we have seen, the correct values on the boundaries. Suppose another solution, giving a value, V', for the potential at any point, were to exist. It is clear that at points on any of the boundary surfaces

$$w = V - V' = 0.$$

Since however

$$\nabla^2 V = 0,$$

and

$$\nabla^2 V' = 0,$$

we must also have

$$\nabla^2 w = 0.$$

Let us substitute the function w for both of the functions represented by U and V in **(3·1)**. We find then

$$\iiint (\textbf{grad } w)^2\, dx\, dy\, dz = \iint w\frac{\partial w}{\partial n}dS,$$

the integration on the left extending over the dielectric, and that on the right over the boundaries. It is clear that the expression on the right is zero, because $w = 0$ at all points on the boundary. The integral on the left is a sum of elements each of which is bound to be either positive or zero; and it can vanish only if

$$\mathbf{grad}\ w = 0$$

everywhere. Therefore

$$w = V - V' = \text{a constant.}$$

In order that w may vanish at the boundaries the constant must be zero. Therefore

$$V - V' = 0$$

everywhere, and the unique character of the solution is established.

§ 20·9. Dielectric Sphere in Uniform Field

We shall now turn to a problem which is related to those with which we have dealt in the last sections. We imagine an infinitely extended isotropic dielectric medium for which the dielectric constant has the value K_1, and at some place within it a dielectric sphere of which the constant is K_2. We shall suppose there are no charges anywhere, and that the field at distant points is uniform, i.e. the field intensity, $\mathcal{E}$, is constant in magnitude and direction at points very distant from the sphere, K_2. The problem before us is the determination of the field intensity at all other points. The solution of the earlier problem of a conducting sphere in a uniform field suggests that we should attempt to get an expression for the field intensity *outside* the dielectric sphere by the device of supposing the sphere removed and its place taken by dielectric material with the constant K_1, a dipole of suitable moment, $\mathbf{M}$, being placed at the centre and directed along $\mathcal{E}$. We shall endeavour to solve the problem by assuming the field *outside* the spherical region to be a superposition of the field due to the dipole on the uniform field, $\mathcal{E}$.

Consider first the tangential component of the intensity, at a point on the spherical boundary, due to $\mathbf{M}$. It is equal to $A\mathbf{M} \sin \theta / K_1 R^3$ by (20·721), where R is the radius of the sphere and θ is shown in Fig. 20·8. This may be written $A\mathbf{M} \sin \phi / K_1 R^3$. The component of $\mathcal{E}$ in the same direction is easily seen to be $- \mathcal{E} \sin \phi$. In the direction shown by the arrow, parallel to the

spherical surface, the total component of the intensity *outside* the sphere is therefore

$$\frac{A\mathbf{M} \sin \phi}{K_1 R^3} - \mathcal{E} \sin \phi. \quad . \quad . \quad . \quad (20 \cdot 9)$$

In order to deal with the field *within* the sphere let us consider any meridian plane, i.e. one containing the diameter which is parallel to the direction of $\mathcal{E}$. The symmetry of the field indicates at once that it has no component perpendicular to this plane. We shall therefore represent the field intensity at any boundary point *within* the sphere as made up of two components : $\mathcal{E}_p{}'$ parallel to $\mathcal{E}$, and $\mathcal{E}_s{}'$ perpendicular to $\mathcal{E}$ and directed towards the diameter; both being of course in the meridian plane. Consequently we find for the tangential component *within* the sphere :

$$- \mathcal{E}_p{}' \sin \phi + \mathcal{E}_s{}' \cos \phi. \quad . \quad . \quad (20 \cdot 91)$$

The condition **(20·32)** requires :

$$\frac{A\mathbf{M} \sin \phi}{R^3} - K_1 \mathcal{E} \sin \phi = -K_1 \mathcal{E}_p{}' \sin \phi + K_1 \mathcal{E}_s{}' \cos \phi. \quad . \quad (20 \cdot 92)$$

We now make use of the condition **(20·35)**. The normal component at a point on the sphere on the *outside* is calculated from $\mathcal{E}$ and $\mathbf{M}$. Along the outward normal we have

$$\frac{2A\mathbf{M} \cos \theta}{K_1 R^3} - \mathcal{E} \cos \phi,$$

or

$$- \frac{2A\mathbf{M} \cos \phi}{K_1 R^3} - \mathcal{E} \cos \phi \; ;$$

and hence for the product of normal component and dielectric constant we get

$$- \frac{2A\mathbf{M} \cos \phi}{R^3} - K_1 \mathcal{E} \cos \phi. \quad . \quad . \quad (20 \cdot 93)$$

For the normal component of the intensity in the interior, in the outward direction, we have

$$- \mathcal{E}_p{}' \cos \phi - \mathcal{E}_s{}' \sin \phi \; ;$$

and hence for the produce of normal component and dielectric constant :

$$- K_2 \mathcal{E}_p{}' \cos \phi - K_2 \mathcal{E}'_s \sin \phi. \quad . \quad . \quad (20 \cdot 94)$$

Equating (20·93) and (20·94), we get

$$\frac{2A\mathbf{M}\cos\phi}{R^3}+K_1\mathcal{E}\cos\phi=K_2\mathcal{E}_p{}'\cos\phi+K_2\mathcal{E}_s{}'\sin\phi. \qquad . \quad (20\cdot95)$$

For the moment we are concerned merely to constitute the field at the boundary in such a way that the boundary conditions expressed by (20·92) and (20·95) are satisfied. We can do this in the simplest way by making $\mathcal{E}_s{}' = 0$. We then get

$$\frac{A\mathbf{M}}{R^3} - K_1\mathcal{E} = - K_1\mathcal{E}', \qquad . \quad . \quad . \quad (20\cdot96)$$

and

$$\frac{2A\mathbf{M}}{R^3} + K_1\mathcal{E} = \quad K_2\mathcal{E}', \qquad . \quad . \quad . \quad (20\cdot97)$$

in which $\mathcal{E}_p{}'$ has been replaced by $\mathcal{E}'$, since it is no longer necessary to distinguish it from $\mathcal{E}_s{}'$. We find at once that

$$\mathcal{E}' = \frac{3K_1}{K_2 + 2K_1}\mathcal{E} ; \qquad . \quad . \quad . \quad . \quad (20\cdot98)$$

that is to say, the field in the interior of the sphere is uniform and parallel to $\mathcal{E}$.

The moment M is found to be :

$$\mathbf{M} = \frac{K_1(K_2 - K_1)R^3}{(K_2 + 2K_1)A}\mathcal{E}. \qquad . \quad . \quad (20\cdot99)$$

It will be noticed that when K_2 is infinite this equation becomes

$$\mathbf{M} = \frac{K_1 R^3 \mathcal{E}}{A}.$$

This is identical with the value for M appropriate for a conducting sphere (20·81) as we should expect. Furthermore, when K_2 is infinite, $\mathcal{E}'$ becomes zero (20·98), as would be the case if the sphere were made of conducting material. Finally, if we make $K_2 = K_1$, we find that $\mathcal{E}'$ becomes equal to $\mathcal{E}$ (20·98), while M becomes zero (20·99), which again is what we expect, since this is equivalent to removing the sphere and filling its place with the original dielectric K_1. These are verifications of the correctness of the solution we have found.

The preceding investigation does not exclude the possibility of other solutions of the problem with which we are dealing ; but it is easy to show that it satisfies not only the boundary conditions (as we have seen), but also all other conditions. The field intensity *at any point outside the sphere* K_2 is a super-

position of the field of the dipole and the constant field, $\mathcal{E}$. It has therefore a potential :

$$V_e = \frac{A\mathbf{M}\cos\theta}{K_1 r^2} + v_e,$$

or $$V_e = \frac{(K_2 - K_1)R^3\mathcal{E}}{(K_2 + 2K_1)r^2}\cos\theta + v_e,$$

where v_e is the potential of the constant field $\mathcal{E}$ and r is the radial distance from the centre of the sphere ; and it is obvious that this satisfies (as it is required to do) Laplace's equation at all points outside the sphere, and that it gives the constant field $\mathcal{E}$ at infinity. The field inside the sphere, being constant in magnitude and direction, also satisfies Laplace's equation.

We can now (following the method of **§ 20·8**) prove that there is no other solution. Let V_e be the potential at all points outside the sphere, given by the solution we have found, and V_i that at all points within the sphere ; and let V_e' and V_i' be the corresponding potentials determined by any other solution, if there be one. Let

$$w_e = V_e - V_e',$$

and $$w_i = V_i - V_i' ;$$

and substitute w_e (or w_i) for both U and V in equation **(3·1)**. We thus obtain, since

$$\nabla^2 w_e = 0,$$

and $$\nabla^2 w_i = 0,$$

$$\left.\begin{array}{l} \iiint \{\mathbf{grad}\ w_e\}^2\ dx\ dy\ dz = \iint w_e\dfrac{\partial w_e}{\partial n}dS, \\[3mm] \iiint \{\mathbf{grad}\ w_i\}^2\ dx\ dy\ dz = \iint w_i\dfrac{\partial w_i}{\partial n}\ dS. \end{array}\right\}\qquad . \quad \textbf{(20·991)}$$

In the former of these two equations the volume integration may be taken to extend over the region outside the sphere, bounded by two infinitely distant planes perpendicular to $\mathcal{E}$, and an infinitely distant cylindrical surface parallel to $\mathcal{E}$. The surface integral is extended over these surfaces and over that of the sphere. In the latter equation the volume integration extends over the volume of the sphere, while the surface integral is extended over its surface. The integration over the infinitely distant surfaces obviously contributes nothing to the total, since $w_e = 0$ at infinity, and we may confine our attention to the spherical surface. The differentiation $\partial/\partial n$ is of course

in the direction of the outward normal, and in the case
of the external region it is equivalent to $- \partial/\partial r$ at the spherical
boundary, and to $+ \partial/\partial r$ in the case of the internal region.
Let us multiply the two equations by K_1 and K_2 respectively
and add. We thus obtain :

$$K_1 \iiint \{\mathbf{grad}\ w_e\}^2\ dx\ dy\ dz + K_2 \iiint \{\mathbf{grad}\ w_i\}^2\ dx\ dy\ dz$$

$$= \iint \left\{ w_i K_2 \frac{\partial w_i}{\partial r} - w_e K_1 \frac{\partial w_e}{\partial r} \right\} dS. \qquad . \qquad . \qquad (20\text{·}992)$$

Now
$$K_1 \frac{\partial w_e}{\partial r} = K_2 \frac{\partial w_i}{\partial r},$$

because the normal component of the electric displacement has
the same value on both sides of the spherical boundary. Further-
more, it follows from (20·31) that $V_e - V_i$ has the same value
at all points on this boundary. So likewise has $V_e' - V_i'$, and
as this constant difference is a property of the boundary,

$$V_e - V_i = V_e' - V_i'$$

or
$$w_i = w_e$$

at the boundary. Consequently the right-hand side of equation
(20·992) is zero. The integrals on the other side consist of a
sum of contributions, each of which, if different from zero at
all, must necessarily be positive. We must therefore conclude
that

$$\mathbf{grad}\ w_e = 0,$$
$$\mathbf{grad}\ w_i = 0.$$

Consequently
$$V_e - V_e' = \text{constant},$$
$$V_i - V_i' = \text{constant} ;$$

which establishes that the assumed alternative solution does not
differ from the original one.

BIBLIOGRAPHY

HENRY CAVENDISH : Electrical Researches.

MICHAEL FARADAY : Experimental Researches.

GEORGE GREEN : An Essay on the Application of Mathematical Analysis
    to the Theories of Electricity and Magnetism.

CLERK MAXWELL : A Treatise on Electricity and Magnetism. (Clarendon
    Press.)

A. G. WEBSTER : Electricity and Magnetism. (Macmillan.)

CLEMENS SCHAEFER : Einführung in die theoretische Physik, Band III.
    (De Gruyter.)

CHAPTER IV

MAGNETOSTATICS

§ 21. Magnetization

THE more obvious phenomena of magnetism are very similar to those of electrostatic fields. Indeed, the quantitative relationships of a large class of magnetic phenomena are *exactly* similar ; though they are nevertheless easily distinguishable. They are most prominently associated with certain elements, especially iron, nickel and others of that class. The **lodestone**—*lode* ≡ *lead* (German, *leiten*)—an oxide of iron (Fe_3O_4), seems to have been discovered, and its salient properties recognized, in early times, in Magnesia in Asia Minor. Hence the origin of the names **magnetite** and **magnetism.** The term **magnet** is a name for any magnetized piece of material, i.e. a piece of material which exerts forces on pieces of iron, steel, nickel, etc., and if suspended by a torsionless fibre will take up a definite geographical orientation when in equilibrium. Pieces of iron, steel, nickel, magnetite, etc., if merely left undisturbed, will gradually become magnetized. Most of the iron retort stands in a laboratory, if tested, will be found to be magnetized.

When a magnetized piece of steel, or other material, is plunged into a mass of iron filings, these are found to adhere in thick masses *to certain parts* of the surface of the steel, and not uniformly all over it. These regions over which the attraction is predominantly exerted are called **poles,** though the stricter usage of the term ' pole ' associates it with a point, generally one of two points on a joining straight line (axis). Every **magnet** (magnetized piece of material) is found to have *two poles or more.*[1] If a magnet with two approximately point poles, i.e. a magnet the forces associated with which are directed to or from two points, be suspended by a torsionless suspension

---

[1] This statement is not strictly correct. A piece of iron or steel in the form of a ring may be magnetized and exhibit no poles whatever. Perhaps one would not term such a magnetized ring a magnet. A piece of material cannot be magnetized in such a way as to exhibit a solitary pole.

(silk fibre) from its centre of gravity, it will be found that it tends to set itself so that the line joining the poles (axis) takes up a definite geographical direction. In the northern hemisphere this direction is, very roughly speaking, through a point on, or near, the geographical axis somewhere above the earth's equatorial plane. The measurements with a *dip circle* represent a more precise form of this experiment. We can thus distinguish between *two kinds* of magnetic poles—**north poles** and **south poles,** a north pole being naturally one which is pulled towards the northern hemisphere. Every magnet is found to have at least one pole of each kind. Very simple experiments indicate that two poles (like two electric charges) repel one another when they are alike, i.e. both north poles or both south poles, and attract one another when they are unlike ; and we may account in a rough and approximate way for the behaviour of the dip circle magnet by assuming a moderately short two-poled magnet (dipole) at the centre of the earth, with its axis coincident with the geographical axis, and its *south* pole directed *northwards*.

By constructing a magnet in the form of a thin steel rod with spherical ends of larger radius, and preferably of soft iron, we can produce approximately point poles. That is to say, the attractions or repulsions exerted by such a magnet on other point poles or on particles of iron (iron filings) are exerted approximately to or from points which are near the centres of the spherical ends. Generally speaking, however, the poles of a magnet are very far from being point poles. It is, however, a convenient (mathematical) device to regard a magnet as constituted of point poles, or more precisely, to regard every volume element in the magnet as containing poles, or, as we shall sometimes say, magnetic charges analogous to point charges in electrostatics ; and we shall define the measure of a **magnetic charge,** or, as we more usually term it, **pole strength,** in a way exactly corresponding to the definition of quantity of electricity in **§ 18·1.** Only we must bear in mind that an isolated **magnetized particle** will contain (at least) *two poles* of opposite kinds and is not therefore analogous to the simplest electrified particle (see **§ 18·1).** There is, however, nothing to hinder us from regarding it (for purposes of calculation) as made up of two parts, one a north pole and the other a south pole. These may be regarded as the analogues of electrified particles. We shall adopt the same type of formula as **(18·1)** to express the force between two such point poles, namely

$$\mathbf{F} = m_1 m_2 \phi(r), \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \textbf{(21)}$$

where m_1 and m_2 are the **pole strengths** of the poles (or their

magnetic charges, as we might term them) and r is the distance between them. The footnote on page 2 applies here with much greater force than it does in electrostatics. The formula (21) represents the facts when the two points poles are in empty space, or in air, or in almost any medium which a rather careless examination would class as non-magnetic.

Without any knowledge of the law of force we can assure ourselves by simple experiments, which need not be described here, that **the algebraic sum of the pole strengths of the poles of a magnet is always zero.** This fact and the further one that **there are no magnetic phenomena corresponding to electric conduction,** sharply distinguish purely magnetic phenomena from purely electrical phenomena. A magnet is the exact analogue of a polarized piece of dielectric material, like the sphere of the preceding section ; provided we assign to the latter one further property which it does not actually possess— at all events not in any appreciable degree—namely that of permanence, that is to say of being able to remain polarized even when the external polarizing field has been removed. Briefly **polarization in a dielectric** resembles very closely **magnetization in a magnet or magnetic material.** Indeed, the definition of **intensity of magnetization** which will presently be given is precisely analogous to that of **polarization** in a dielectric medium. We shall define **magnetic field intensity** in similar terms to those used in § 18·2 to define electric field intensity ; the only difference being that we are here concerned with force per unit north pole. We shall usually represent it by the symbol, **H.** Similarly, we shall introduce the conception of a magnetic displacement, $\mathbf{D}_m$, defined by

$$\mathbf{D}_m = \frac{\mu \mathbf{H}}{4\pi A}, \qquad . \quad . \quad . \quad . \quad \textbf{(21·01)}$$

in which μ, the **permeability** of the medium, is a quantity analogous to the dielectric constant ; though for many media, including iron and nickel, it is not a constant. The **intensity of magnetization** in a medium we shall define by

$$\mathbf{I} = \frac{\mu - \mu_0}{4\pi A}\mathbf{H}, \qquad . \quad . \quad . \quad . \quad \textbf{(21·02)}$$

a formula which is analogous to **(19·411).**

Considerations analogous to those discussed in § 19·2 lead to

$$\text{Div } \mathbf{D}_m = \rho_m,$$

where ρ_m means density of magnetic charge. We have seen, however, that this is always zero ; hence

$$\text{Div } \mathbf{D}_m = 0. \qquad . \quad . \quad . \quad . \quad \textbf{(21·03)}$$

§ 21·1. Magnetic Moment

By the **magnetic moment** of a magnet we understand a vector $\mathbf{M} \equiv (M_x, M_y, M_z)$, such that

$$M_x = \Sigma m_s x_s. \quad . \quad . \quad . \quad . \quad \textbf{(21·1)}$$

This equation is to be interpreted as follows : we regard the magnet as an assemblage of point poles 1, 2, 3 . . . s, m_s being the strength of the pole, s, and x_s its X co-ordinate (Fig. 21·1). The Y and Z components of $\mathbf{M}$ are similarly defined.

If a magnet be placed in a field of uniform intensity, $\mathbf{H}$, the resultant force exerted on it is found to be zero, and since the force is

Fig. 21·1

$$\mathbf{F} = \Sigma m_s \mathbf{H} = \mathbf{H} \Sigma m_s,$$

it is evident that

$$\Sigma m_s = 0, \quad . \quad . \quad . \quad . \quad \textbf{(21·11)}$$

a result to which we have already alluded (§ 21).

The resultant couple exerted on a magnet placed in a uniform field $\mathbf{H}$ will be made up of the moments of the forces exerted on the individual point poles of which the magnet may be regarded as constituted. We have for the sum of the X components of these moments reckoned with respect to the origin (see § 5·43)

$$\Sigma(y_s m_s H_z - z_s m_s H_y) = M_y H_z - M_z H_y.$$

So that the resultant couple is

$$[\mathbf{M}, \mathbf{H}]. \quad . \quad . \quad . \quad . \quad \textbf{(21·12)}$$

That is to say, it is at right angles to $\mathbf{M}$ and $\mathbf{H}$, in the direction in which an ordinary screw would travel if rotated in the sense from $\mathbf{M}$ to $\mathbf{H}$; while its numerical value is $MH \sin \theta$, where θ is the angle between the directions of $\mathbf{M}$ and $\mathbf{H}$. The work done in producing a small increment $d\theta$ in the angle θ is obviously

$$MH \sin \theta \, d\theta$$

or $$- d\{MH \cos \theta\}.$$

Hence, apart from an arbitrary constant of integration, the
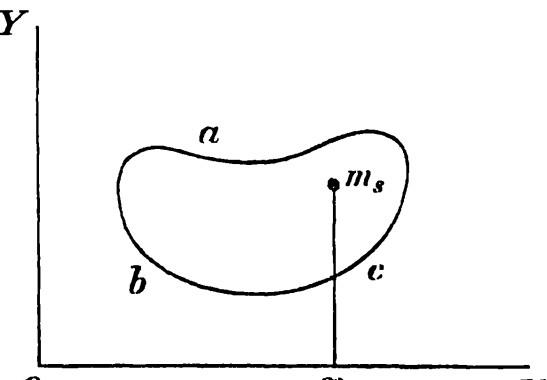
potential energy of a magnet in an external field of intensity **H** is

$$- MH \cos \theta, \quad . \quad . \quad . \quad . \quad . \quad \textbf{(21·13)}$$

θ being the angle between the directions of **M** and **H** or, as we often express it, the angle between the directions of the **magnetic axis** of the magnet and **H**. The **magnetic axis** of a magnet is any straight line which has the same direction as its magnetic moment.

§ 21·2.   THE MAGNETIC POTENTIAL AT A POINT IN THE FIELD OF A SMALL MAGNET

We are now going to study the potential, due to a magnet of any shape or type of magnetization, at a point outside the magnet in air, or preferably *in vacuo*, and very distant from any point on or in the magnet itself. That is to say, if **r** (Fig. 21·2) be

the distance of the point, P, where the potential, V, is to be determined, from some arbitrarily chosen point within the magnet, then **r** is very great when compared with the linear dimensions of the magnet. This explains the sense of the description *small magnet*

FIG. 21·2
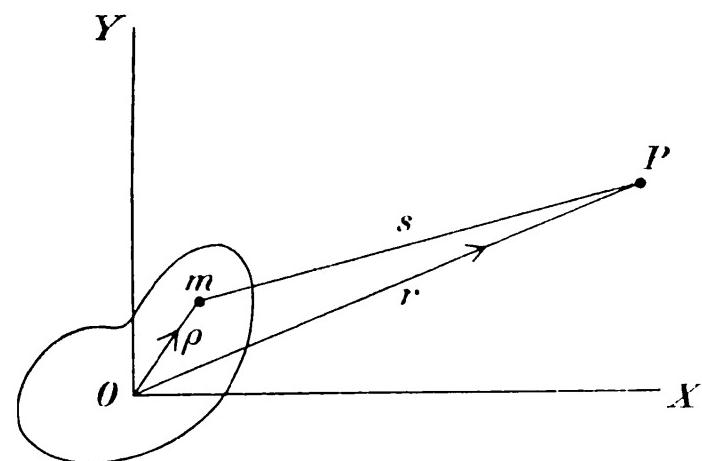
used above. We shall suppose the fixed point, from which **r** is measured, to be the origin of rectangular co-ordinates (Fig. 21·2). Let m be the pole strength of any small element of volume in the magnet, and let us represent its distance from the origin by **ρ**. We shall treat **r** and **ρ** as vectors directed away from the origin. Let s be the distance from m to P. The contribution of m to the potential at P may be written in the form :

$$m\psi(s),$$

where $\psi(s)$ has the property :

$$\phi(s) = - \frac{d}{ds}\psi(s).$$

Now
$$s^2 = (x - \xi)^2 + (y - \eta)^2 + (z - \zeta)^2,$$

where (x, y, z) and (ξ, η, ζ) are the co-ordinates of P and m respectively. Therefore

$$s^2 = x^2 + y^2 + z^2 - 2(x\xi + y\eta + z\zeta),$$

or
$$s^2 = \mathbf{r}^2 \left\{ 1 - \frac{2(x\xi + y\eta + z\zeta)}{\mathbf{r}^2} \right\},$$

if we neglect small quantities of the second order. To the same degree of approximation we have

$$s = \mathbf{r} \left\{ 1 - \frac{(x\xi + y\eta + z\zeta)}{\mathbf{r}^2} \right\}$$

or finally,

$$s = \mathbf{r} - \frac{(\mathbf{r}, \boldsymbol{\rho})}{\mathbf{r}}, \quad . \quad . \quad . \quad . \quad . \quad \textbf{(21·2)}$$

where $(\mathbf{r}, \boldsymbol{\rho})$ is the scalar product of the vectors $\mathbf{r}$ and $\boldsymbol{\rho}$.

By Taylor's theorem the function $\psi(s)$ is therefore

$$\psi(s) = \psi(\mathbf{r}) - \frac{(\mathbf{r}\boldsymbol{\rho})}{\mathbf{r}} \frac{d\psi(\mathbf{r})}{d\mathbf{r}},$$

or

$$\psi(s) = \psi(\mathbf{r}) + \frac{(\mathbf{r}\boldsymbol{\rho})}{\mathbf{r}} \phi(\mathbf{r}), \quad . \quad . \quad . \quad \textbf{(21·21)}$$

neglecting second order small quantities.

In the summation $\Sigma m \psi(s)$, the first term of **(21·21)** contributes nothing, since $\psi(r)$ is a constant and $\Sigma m = 0$ by **(21·11)**. Therefore, if V represent the potential at P,

$$V = \Sigma m \frac{(\mathbf{r}\boldsymbol{\rho})}{\mathbf{r}} \phi(\mathbf{r}),$$

whence
$$V = \frac{\phi(\mathbf{r})}{\mathbf{r}} \{ x\Sigma m\xi + y\Sigma m\eta + z\Sigma m\zeta \},$$

or
$$V = \frac{\phi(\mathbf{r})}{\mathbf{r}} \{ xM_x + yM_y + zM_z \}. \quad . \quad . \quad \textbf{(21·22)}$$

For the X component of the magnetic field intensity at P we have of course

$$f_x = -\frac{\partial V}{\partial x},$$

or, if we remember that $\partial r / \partial x = x/\mathbf{r}$, and if we denote $\dfrac{d\phi(\mathbf{r})}{d\mathbf{r}}$ by $\phi'(\mathbf{r})$,

$$f_x = -\frac{x}{\mathbf{r}^2} \phi'(\mathbf{r}) \{ xM_x + yM_y + zM_z \}$$

$$+ \frac{x}{\mathbf{r}^3} \phi(\mathbf{r}) \{ xM_x + yM_y + zM_z \}$$

$$- \frac{\phi(\mathbf{r})}{\mathbf{r}} M_x. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad \textbf{(21·23)}$$

Of special interest is the component of the field intensity in a direction parallel to the axis of the magnet. We shall therefore imagine the X axis of the co-ordinate system to be turned into this direction. In consequence

$$M_x = \mathbf{M},$$
$$M_y = 0,$$
$$M_z = 0,$$

and (21·23) becomes

$$f_x = -\frac{x^2}{\mathbf{r}^2}\phi'(\mathbf{r})\mathbf{M} + \frac{x^2}{\mathbf{r}^3}\phi(\mathbf{r})\mathbf{M} - \frac{\phi(\mathbf{r})}{\mathbf{r}}\mathbf{M}. \qquad . \quad (21\cdot24)$$

If the point P be *on the axis through* O,

$$x = \mathbf{r},$$
$$y = 0,$$
$$z = 0,$$

and (21·24) becomes

$$f_x = {}_1f_x = -\phi'(\mathbf{r}).\mathbf{M}, \qquad . \quad . \quad . \quad (21\cdot25)$$

while, if P be on the Y axis, for example,

$$x = 0,$$
$$y = \mathbf{r},$$
$$z = 0,$$

and we get from (21·24) :

$$f_x = {}_2f_x = -\frac{\phi(\mathbf{r})}{\mathbf{r}}\mathbf{M}. \quad . \quad . \quad . \quad (21\cdot26)$$

Experiment indicates that

$$\frac{{}_1f_x}{{}_2f_x} = -2,$$

for the same value of $\mathbf{r}$ in the two cases ; therefore

$$\mathbf{r}\frac{\phi'(\mathbf{r})}{\phi(\mathbf{r})} = -2,$$

or

$$\frac{\phi'}{\phi} + \frac{2}{\mathbf{r}} = 0,$$

where $\phi(\mathbf{r})$ and $\phi'(\mathbf{r})$ are abbreviated to ϕ and ϕ'. Hence

$$\frac{d\log\phi}{d\mathbf{r}} + 2\frac{d\log\mathbf{r}}{d\mathbf{r}} = 0,$$

and consequently

$$\mathbf{r}^2\phi = \text{constant},$$

that is to say,

$$\phi(\mathbf{r}) = \frac{\alpha}{\mathbf{r}^2}, \qquad . \quad . \quad . \quad . \quad . \quad (21\cdot27)$$

where α is a constant, analogous to the constant α of § 18·1.

This result expresses the law of force between point poles separated by a medium like empty space, air or anything which possesses very feeble magnetic properties. It is convenient to express α in a form similar to (19·08). Thus

$$\phi(r) = \frac{A}{\mu r^2}, \qquad . \quad . \quad . \quad . \quad . \quad (21\cdot28)$$

where again A is pure number depending on our choice of units, and is usually either unity or $1/4\pi$.

If we substitute (21·28) in (21·22), we get

$$V = \frac{A}{\mu \mathbf{r}^3}\{xM_x + yM_y + zM_z\}$$

or

$$V = \frac{A(\mathbf{r}, \mathbf{M})}{\mu \mathbf{r}^3}. \qquad . \quad . \quad . \quad . \quad . \quad (21\cdot29)$$

This is of course equivalent to

$$V = \frac{A\mathbf{M}\cos\theta}{\mu \mathbf{r}^2}, \qquad . \quad . \quad . \quad . \quad (21\cdot291)$$

where θ is the angle between the direction of $\mathbf{r}$ and that of the axis of the magnet. This formula should be compared with (20·71).

CHAPTER V

FOUNDATIONS OF ELECTROMAGNETISM

§ 21·3. ELECTRIC CURRENTS

EXPERIMENT shows that when a charged conductor, or condenser, is discharged through a coil of wire a magnetic field is produced in the surrounding neighbourhood. We shall define the **strength** of an **electric current** (in a conductor such as a piece of wire, for example) as the quantity of electricity passing any cross-section of the wire per unit time. To make the notion of an electric current, and of current strength and direction, as clear as possible, we may note that a transference of positive electricity—to fix our ideas, let us say a transference of + 7 units—from a conductor A to another B produces charges on A and B which are identical with those due to a numerically equal transference of negative electricity from B to A, as we have already learned in the sections devoted to electrostatics. Similarly, we cannot distinguish the magnetic effect of a current of + 7 units per second transferred in *one* direction from that of a current of − 7 units per second transferred in the opposite direction. When we speak of the **direction of a current** we mean that direction in which positive electricity would have to travel to produce the magnetic or other effects of the current.

The current strength in a wire *may vary from one cross-section to another* ; but we shall usually be concerned with cases where this variation does not occur, or is negligible. Electrostatic devices are occasionally used for measuring currents, as, for example, in the measurement of ionization currents. That is to say, the quantity of electricity passing some cross-section in a measured time is determined from the rise of potential of some system of known capacity. More usually current strengths are determined from measurements of the intensity at some point of the associated magnetic field.

Just as in § **19·2** the quantity of electricity displaced through the vector element $\mathbf{dS}$, in the sense of its vector arrow, is $(\mathbf{D}, \mathbf{dS})$,

so the quantity passing per unit time through **dS** in this sense may be written

$$(\mathbf{i}, \mathbf{dS}). \quad . \quad . \quad . \quad . \quad . \quad (21\cdot3)$$

We shall term the vector **i** the **current density** in the neighbourhood of **dS**. A study of the mechanical picture of § 19·2 leads at once to the conclusion,

$$\frac{\partial i_x}{\partial x} + \frac{\partial i_y}{\partial y} + \frac{\partial i_z}{\partial z} = 0, \quad . \quad . \quad . \quad (21\cdot31)$$

or by the theorem of Gauss,

$$\iint (\mathbf{i}\, \mathbf{dS}) = 0, \quad . \quad . \quad . \quad . \quad (21\cdot32)$$

where the integration is extended over any closed surface. This equation states that the algebraic or net quantity of electricity emerging per unit time from any closed surface is equal to zero. We shall apply this result to the case where a conduction current is flowing along a wire, and where a steady state exists, i.e. such a state that the field at all points remains constant. It is clear that, in such a case, the displacement currents in the surrounding dielectric must be everywhere zero, and consequently equation (21·32) necessitates that the current strength has the same value at all cross-sections of the wire.

§ 21·4. MAINTENANCE OF CURRENTS

When a current is maintained in a closed conducting circuit (*ABC* in Fig. 21·4), it is clear that the associated phenomena differ from those of an electrostatic field, since we have here electricity in motion and an associated magnetic field in the neighbourhood of the circuit. But even when a steady state exists, and the electric field intensity at all points remains constant, and in this broad sense may be described as *static*, it nevertheless differs in one very important respect from an electrostatic field. Since the electric intensity at every point on the circuit must have a positive

Fig. 21·4

component in the direction of flow of electricity (see Fig. 21·4), it is clear that the line integral
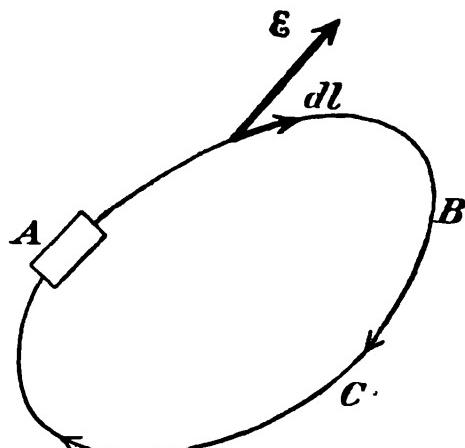
$$\oint (\boldsymbol{\varepsilon}, \mathbf{dl}). \quad . \quad . \quad . \quad . \quad (21\cdot4)$$

round the circuit cannot vanish. It therefore follows that the electric intensity, $\mathcal{E}$, cannot *everywhere* be expressed by

$$\mathcal{E} = - \text{grad } V, \quad . \quad . \quad . \quad . \quad (21 \cdot 41)$$

as would be the case in an electrostatic field. This does not of course mean that such a function as V does not exist. It means that in general there are other factors contributing to $\mathcal{E}$ besides the negative gradient of the **scalar potential** (as we may term it ; since the term ' electrostatic ' is now rather out of place). Our general expression for $\mathcal{E}$ must therefore take the form

$$\mathcal{E} = \mathcal{E}' - \text{grad } V, \quad . \quad . \quad . \quad (21 \cdot 42)$$

where $\mathcal{E}'$ is the part of the electric field intensity not derivable from a scalar potential. It may happen that $\mathcal{E}'$ only differs from **zero** in some restricted region such as A in Fig. 21·4. A familiar example of this is furnished by the case where a steady current is being generated by a cell of some kind. Inside the cell we have

$$\mathcal{E} = \mathcal{E}' - \text{grad } V,$$

$\mathcal{E}'$ and $-$ grad V being approximately in opposite directions. Outside the cell we have $\mathcal{E} = -$ grad V. When the wire connecting the poles of the cell is cut, or removed, the current through the cell sinks to zero, and we have in consequence everywhere within it :

$$\mathcal{E}' - \text{grad } V = 0 ; \quad . \quad . \quad . \quad (21 \cdot 43)$$

since the interior of a cell is a conductor and zero current is only compatible with zero intensity.

In order to maintain a current in any circuit a continuous expenditure of energy is necessary. When the current is increasing, for example, magnetic and electric fields are being built up in the surrounding neighbourhood, the energy per unit volume in the latter case being **(§ 19·9)** $\frac{1}{2}(\mathcal{E}, \mathbf{D})$, where $\mathcal{E}$ is the electric intensity and $\mathbf{D}$ the associated displacement. A similar formula expresses the energy per unit volume in the magnetic field. Even when the current is steady an expenditure of energy is necessary, since heat is produced in the circuit. In the case where the current is generated by a cell, the energy is supplied through the agency of the chemical reactions occurring in it ; in the case of a thermo-circuit energy in the form of heat is supplied or abstracted at the junctions and in other parts of the circuit ; and a current may be maintained by the excess of the heat absorbed over that emitted.

§ 21·5. Electromotive Force

The quantity of energy supplied to a circuit, per unit quantity of electricity passing round it, is called the **electromotive force** in the circuit. Since the intensity, $\mathcal{E}$, means the force per unit quantity, the work done in the displacement of the unit quantity through the distance **dl** is of course $(\mathcal{E}, \mathbf{dl})$, and consequently if E be the electromotive force,

$$\mathrm{E} = \oint (\mathcal{E}, \mathbf{dl}),$$

where the symbol $\oint$ indicates that the integration is extended round the whole circuit. Therefore

$$\mathrm{E} = \oint (\mathcal{E}', \mathbf{dl}) - \oint (\mathrm{grad}\ V, \mathbf{dl}),$$

and consequently

$$\mathrm{E} = \oint (\mathcal{E}', \mathbf{dl}),\ .\quad .\quad .\quad .\quad .\quad \mathbf{(21·5)}$$

since, by Stokes's theorem **(3·32)**,

$$\oint (\mathrm{grad}\ V, \mathbf{dl}) = 0.$$

Turning to the simple case where the current is due to a single cell, the integral **(21·5)** is termed the electromotive force of the cell. When the current is a steady one (not varying), $\mathcal{E}'$ differs from zero *within the cell only*, and the electromotive force of the cell is the integral $\int (\mathcal{E}'\mathbf{dl})$ taken through the cell from the negative pole to the positive pole. When the cell is on open circuit, in other words, when the wire connecting its poles is broken, we have by **(21·43)**

$$\int (\mathcal{E}', \mathbf{dl}) = \int (\mathrm{grad}\ V, \mathbf{dl}),$$

the integration extending from the negative to the positive pole. Therefore if E be the electromotive force of the cell,

$$\mathrm{E} = \int_{-\text{pole}}^{+\text{pole}} \left(\frac{\partial V}{\partial x} dx + \frac{\partial V}{\partial y} dy + \frac{\partial V}{\partial z} dz \right),$$

or

$$\mathrm{E} = \int_{-\text{pole}}^{+\text{pole}} dV = V_+ - V_-,$$

where V_+ and V_- are the potentials of the positive and negative poles respectively. We may therefore say that the electromotive force of the cell is equal to the potential difference between its poles when no current is flowing through it. It will be noticed that, while the integral $\int (\mathcal{E}', d\mathbf{l})$, taken through the cell, means the supply of energy, reckoned per unit quantity of electricity passing through it, *this integral is* **not** *equal to the work done on the unit quantity* **within** *the cell.* This latter quantity is equal to

$$\int (\mathcal{E}d\mathbf{l}) = \int (\{\mathcal{E}' - \text{grad } V\}, d\mathbf{l}),$$

the difference, namely

$$\int (\text{grad } V, d\mathbf{l}),$$

being done on the unit quantity **outside** the cell.

When a current is maintained in a circuit by several current generating agencies, e.g. by a number of cells, the electromotive force of each cell—we shall usually denote it briefly by **E.M.F.**—in any direction is represented by

$$\text{E.M.F.} = \int (\mathcal{E}', d\mathbf{l}), \quad . \quad . \quad . \quad . \quad \textbf{(21·51)}$$

the integration being extended through the cell in that direction.

§ 21·6. Electrical Resistance and Conductance

The work that has to be done to maintain a current in a circuit will depend on various factors. It may be done in building up a magnetic field, in driving a motor which may be in the circuit, and in other ways. The simplest case is that of a steady current in a conducting circuit ; the work is done exclusively in generating heat in the circuit, and the circumstances are precisely analogous to those where work is done in making one body slide over another with a constant velocity. In such a case the work is done in overcoming the frictional resistance, with a consequent development of heat. We are thus naturally led to the conception of **electrical resistance.** We shall take as a measure of the electrical resistance of a conductor (a piece of wire for example) the value of $\int (\mathcal{E}, d\mathbf{l})$

along the wire when the unit current is flowing through it, and the resistance, R, is therefore

$$R = \frac{\int (\mathcal{E}, d\mathbf{l})}{i}. \qquad \ldots \quad (21\cdot6)$$

We are supposing, of course, that the work is done in overcoming a hindrance due to the inherent properties of the conducting material, with a consequent development of heat in the conductor; that it is not done, for example, in contributing to the energy of a magnetic or electric field, or in driving a motor. The reciprocal of the resistance of a conductor is called its **conductance.**

We have then

$$\int (\mathcal{E}, d\mathbf{l}) = Ri$$

or

$$\int (\mathcal{E}', d\mathbf{l}) - \int (\mathrm{grad}\ V, d\mathbf{l}) = Ri,^1$$

and consequently, if we number the terminal points 1 and 2 and represent the E.M.F. and the current in the direction 1 to 2 by $\mathbf{E}_{12}$ and i_{12} respectively,

$$\mathbf{E}_{12} + V_1 - V_2 = R_{12}i_{12}. \qquad \ldots \quad (21\cdot61)$$

The resistance of a conductor is always independent of the direction of flow of the current, so that $R_{12} = R_{21}$.

§ 21·7.   Laws of Ohm and Kirchhoff

It is an experimental fact that in many conductors, e.g. in all solid and liquid metallic or electrolytic conductors, R is independent of the current strength, though it depends on the temperature of the conductor, and in some cases on other factors as well. The independence of R of the current strength is known as **Ohm's law.** Gases and vapours furnish conspicuous exceptions to Ohm's law.

Problems concerning the distribution of current and potential in a network of conductors are most conveniently solved by the use of the relationships usually called **Kirchhoff's laws.** One of these we have met already. It asserts that the algebraic sum of the strengths of the currents flowing into (or away from)

---

¹ We are assuming that the current, i, is a steady current, i.e. that it remains constant.

any point of the network (the point 2 in Fig. 21·7 (a) for example) is equal to zero. The remaining law of Kirchhoff can be deduced from (21·61). It asserts that the E.M.F. round any mesh of the network (the mesh 1, 2, 3 in Fig. 21·7 (a) for example) is equal to

$$R_{12}i_{12} + R_{23}i_{23} + R_{31}i_{31},$$

or briefly

$$\text{E.M.F.} = \Sigma R i. \qquad \qquad (21\cdot7)$$

We infer from (21·61) :

$$\left. \begin{array}{l} E_{12} + V_1 - V_2 = R_{12}i_{12}, \\ E_{23} + V_2 - V_3 = R_{23}i_{23}, \\ E_{31} + V_3 - V_1 = R_{31}i_{31}, \end{array} \right\} \qquad . \quad . \quad . \quad (21\cdot71)$$

and Kirchhoff's law is seen to be an immediate consequence of adding these equations.
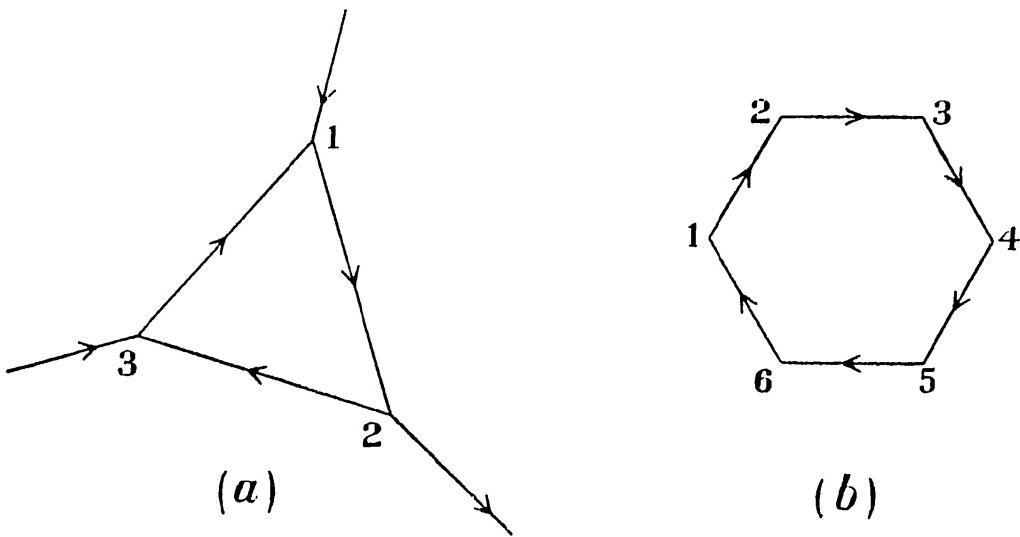
Fig. 21·7

Imagine a circuit (not part of a larger network) such as that illustrated in Fig. 21·7 (b). Let us suppose the E.M.F.'s in the portions 12, 23, 34, etc., all equal to one another. That is to say,

$$E_{12} = E_{23} = E_{34} = \ldots = E_{n1} = E,$$

if we think of the circuit as built up of n such portions. Clearly in this case

$$i_{12} = i_{23} = i_{34} = \ldots = i_{n1} = i.$$

Then we have :

$$E + V_1 - V_2 = R_{12}i$$
$$E + V_2 - V_3 = R_{23}i$$
$$E + V_3 - V_4 = R_{34}i$$
$$\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot$$
$$E + V_n - V_1 = R_{n1}i.$$

$$\left. \right\} \qquad . \quad . \quad . \quad (21 \cdot 72)$$

Consequently

$$nE = i(R_{12} + R_{23} + \ldots + R_{n1}),$$

or

$$E = i\frac{R_{12} + R_{23} + \ldots R_{n1}}{n},$$

and therefore

$$E = i\overline{R},$$

where $\overline{R}$ is the average resistance of the portions 12, 23, 34, ... $n1$, of the circuit. Substitute this result for E in any one of the equations (21·72) : in the 34 equation for example, and we get

$$i\overline{R} + V_3 - V_4 = iR_{34}.$$

If now the resistance R_{34} happens to be equal to the average resistance $\overline{R}$, we must have

$$V_3 - V_4 = 0,$$

or

$$V_3 = V_4 ;$$

and indeed if we suppose the resistance, R, and the electromotive force, E, in the branch, 34, to be uniformly distributed along it, that is to say if the component of $\mathcal{E}'$ in the direction of flow of the current is constant along 34, it is easy to show the potential has the same value at all points along it.

§ 21·8. ELECTRICAL CONDUCTIVITY AND RESISTIVITY

If in (21·6) R represents the resistance of a piece of material of length dl measured in the direction of the current, and if we use the letter i to represent the current *per unit area* (**current density**), instead of representing as in (21·6) the total current, the formula becomes

$$R = \frac{\mathcal{E}_p dl}{idS},$$

where dS is the cross-sectional area of the conductor and $\mathcal{E}_p$ is the component of the field intensity in the direction dl. This formula may be written

$$\frac{RdS}{dl} = \frac{\mathcal{E}_p}{i}.$$

The quantity

$$\rho = \frac{R dS}{dl} \qquad . \quad . \quad . \quad . \quad . \quad \textbf{(21·8)}$$

is called the **resistivity** of the material, and its reciprocal,

$$\sigma = 1/\rho, \quad . \quad . \quad . \quad . \quad . \quad \textbf{(21·81)}$$

is called its **conductivity**; so that

$$\left. \begin{array}{l} \mathcal{E}_p = \rho i, \\ i = \sigma \mathcal{E}_p. \end{array} \right\} \qquad . \quad . \quad . \quad . \quad \textbf{(21·82)}$$

and

§ 21·9. Joule's Law

The work done in maintaining a current against the resistance of a conductor (a wire) is evidently

$$i \int (\mathcal{E} d\mathbf{l})$$

per unit time, where i means the *total* current in the wire (not the current density), since $\int (\mathcal{E} d\mathbf{l})$ is the work done per unit quantity. And in this case

$$\int (\mathcal{E} d\mathbf{l}) = Ri \; ;$$

consequently the rate of production of heat is

$$Ri^2 \qquad . \quad . \quad . \quad . \quad . \quad . \quad \textbf{(21·9)}$$

or

$$i^2/K, \quad . \quad . \quad . \quad . \quad . \quad \textbf{(21·91)}$$

when K represents the conductance of the wire. Consequently, when R (or K) is constant, as it will be when Ohm's law holds and the physical state of the conductor (its temperature in particular) is not allowed to vary, the rate of generation of heat is proportional to the square of the current strength. This is usually called **Joule's law**. If we consider a small piece of conducting material of length dl in the direction of $\mathcal{E}$, or of the current, and of cross-sectional area dS, we have from **(21·9)** for the rate of generation of heat

$$R dS \; i^2 dS,$$

where i now means current density. On dividing by the volume,

$dSdl$, we get for the rate of generation of heat per unit volume
the expression

$$\frac{RdS}{dl}i^2$$

or

$$\rho i^2, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad \textbf{(21·92)}$$

which may of course be written

$$i^2/\sigma. \quad . \quad . \quad . \quad . \quad . \quad \textbf{(21·93)}$$

§ 22. Magnetic Field of a Small Circuit

It is an experimental fact that when the current strength
in a small circuit, or its dimensions, are suitably adjusted, as
also the orientation of the circuit, the consequent magnetic field
can be made exactly like that of
a given small magnet. By 'small
circuit' and 'small magnet' is
meant a circuit, or magnet, the
linear dimensions of which are small
compared with the distances (from
the circuit or magnet) of the points
in the magnetic field with which
we concern ourselves. When the
circuit consists of a plane loop of
wire (Fig. 22), the direction of the
magnetic axis of the equivalent
magnet is perpendicular to the
plane of the loop and is related to
the direction of the current in

Fig. 22

precisely the same way as the direction of travel of an ordinary
screw is related to the direction in which it rotates.   Experiment
shows that the magnetic field intensity at any point is propor-
tional to the product of the current strength and the area of
the loop.   The formulae **(21·29)** and **(21·291)** therefore become,
when applied to the field (at distant points) of the circuit :
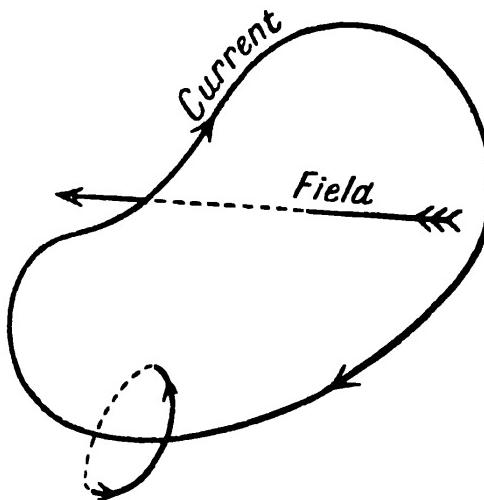
$$V = \beta\frac{i(\mathbf{r}, \mathbf{dS})}{\mathbf{r}^3} \quad . \quad . \quad . \quad . \quad . \quad \textbf{(22)}$$

and

$$V = \beta\frac{i\mathbf{dS}\cos\theta}{\mathbf{r}^2}. \quad . \quad . \quad . \quad \textbf{(22·01)}$$

In these expressions the vector $\mathbf{dS}$ is numerically equal to the
area of the small loop and is assigned the direction of the magnetic

7

axis. The positive constant β has the same value for all media, and is therefore a pure number. It is convenient to replace β by A/a, where A is the same numerical constant (usually 1 or $1/4\pi$) as that represented by A in the formulae of electrostatics and magnetostatics, while a is a new numerical constant. We shall consequently write (**22·01**) in the following way :

$$V = \frac{Ai\mathbf{dS}\cos\theta}{a\mathbf{r}^2}. \qquad . \quad . \quad . \quad (22\cdot02)$$

§ 22·1. Systems of Units

The constant A is always, or nearly always, chosen to have one or other of the values 1 or $1/4\pi$. In the former case we shall describe the units used as **ordinary units** ; in the latter case as **Lorentz-Heaviside units.** The usual choice for the constant a is either 1 or, alternatively, the number which expresses, as we shall see later, the velocity of electromagnetic waves in empty space. It must be emphasized, however, that a is a pure number and not a quantity with the dimensions of a velocity. We shall use the letter c to represent the velocity of electromagnetic waves in empty space. The following table exhibits clearly various systems of **absolute units** which are in use :

$$a = 1 \begin{cases} \begin{aligned} &K = 1 \\ &\text{in empty space} \end{aligned} \begin{cases} A = 1 & \text{Ordinary electrostatic.} \\ A = 1/4\pi & \text{Lorentz-Heaviside electro-} \\ & \qquad\qquad\qquad\text{static.} \end{cases} \\ \begin{aligned} &\mu = 1 \\ &\text{in empty space} \end{aligned} \begin{cases} A = 1 & \text{Ordinary electromagnetic} \\ & \qquad\qquad\qquad\qquad\text{units.} \\ A = 1/4\pi & \text{Lorentz-Heaviside electro-} \\ & \qquad\qquad\qquad\text{magnetic units.} \end{cases} \end{cases}$$

As we shall learn in **§ 24·8,** the velocity of electromagnetic waves in an isotropic medium of dielectric constant, K, and permeability μ is

$$\frac{a}{\sqrt{\mu K|}}.$$

Consequently we have

$$c = \frac{a}{\sqrt{\mu_0 K_0|}}, \qquad . \quad . \quad . \quad . \quad (22\cdot1)$$

where μ_0 and K_0 are respectively the permeability and dielectric constant of empty space. If therefore we choose a definite numerical value for a we cannot *arbitrarily* fix *both* of the units

for K and μ. For instance, when we adopt electrostatic units, whether of the ordinary or of the Lorentz-Heaviside type, we make $a = 1$ (for all media of course, since it is merely a number) and $K = 1$ for empty space. The unit for μ is thereby fixed. It is the unit which makes μ for empty space equal to $1/c^2$, in consequence of (22·1). Similarly, when we adopt electromagnetic units we give a the universal value 1 and make $\mu = 1$ for empty space. In consequence we find for empty space

$$K = \frac{1}{c^2}.$$

If we choose the numerical value of a so that $a = c$ we may then, as formula (22·1) indicates, choose units for K and μ, so that both

$$\left.\begin{array}{l} K = 1 \\ \mu = 1 \end{array}\right\} \text{for empty space.}$$

and

When our units are chosen in this way we shall call them **mixed units**, since they will be found to express the electrical quantities in terms of electrostatic units, and the magnetic quantities in electromagnetic units. Whether they are of the *ordinary* or the *Lorentz-Heaviside* type will depend, of course, solely on the choice made for A. Reference will be made later (§ 24·3) to the units usually known as **practical units**.

§ 22·2. MAGNETIC FIELD OF A CURRENT

In order to approach the general problem of the field of a current let us imagine a loop of thin wire round which a current of strength i is flowing, as indicated in Fig. 22·2. The portion of the loop repre-
sented by a broken line may be sup-
posed to be behind the plane of the paper. Let P repre-
sent any point. It may be conveni-
ently thought of as in the plane of the paper. It will be seen that the mag-

Fig. 22·2

net, to which the current is equivalent, will have its north pole side towards P, and a point north pole at P will experience a force tending to drive it farther away from the current circuit.
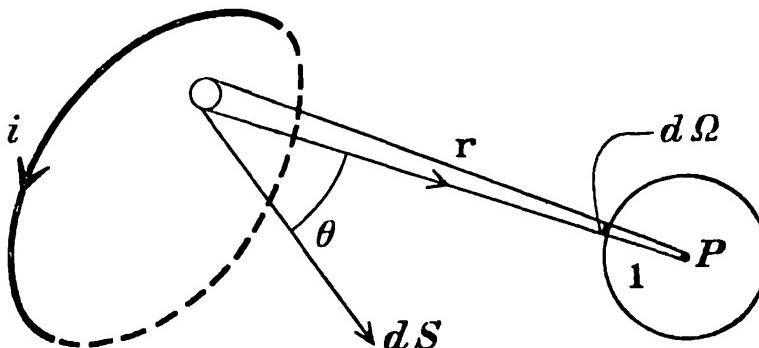
If P be very distant from the latter the formula (22·02) will suffice to describe the field near P, if dS be replaced by the equivalent area. We shall, however, still be able to make use of this formula even when P is not a distant point. Let us imagine a thin sheet or surface having the loop of wire as its boundary. This sheet may have any shape subject to the condition about its boundary. Now suppose a network to be traced on the sheet, so as to divide it into a very large number of small surface elements dS, as shown in Fig. 22·21 (*a*). Instead of the original current of strength i flowing round the boundary of the sheet as shown in Fig. 22·2, let us imagine a large number of minute circuits, the current in each flowing in the same sense round the boundary of a surface element dS. Every element dS, we shall suppose, has such a current flowing round it, and we may take the linear dimensions of every dS to be infinitesimal by comparison with the distance between it and the point P. Such a system of small circuits will give rise to a magnetic field identical with that due to the original circuit consisting of the current i flowing round the boundary loop. This will easily be understood by studying two contiguous elementary circuits as illustrated by Fig. 22·21 (*b*). Equal and oppositely directed currents, it will be seen, flow along their common boundary, so that the current strength in any boundary between two or more such elements is zero. The only current not annulled in this way is clearly that flowing round the loop bounding the sheet.
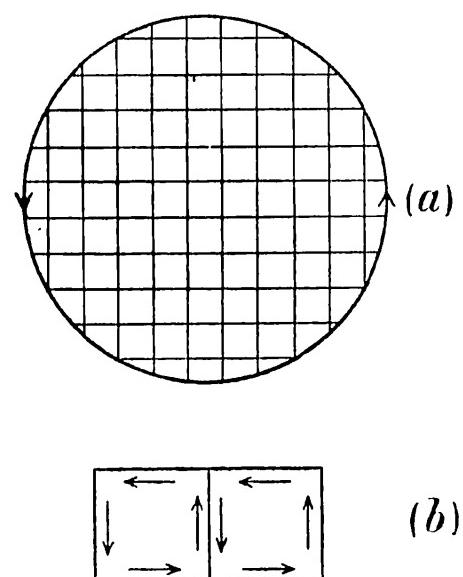
FIG. 22·21

In applying (22·02) to any one of the elements dS, we shall lay down that the vector arrow of dS is directed in the sense of the screw rule (see the last paragraph of § 3·3), so that in Fig. 22·2 it is directed to the same side of the sheet as that on which P is situated. It is now evident that

$$\frac{\mathbf{dS} \cos \theta}{\mathbf{r}^2} = d\Omega,$$

where $d\Omega$ is the small solid angle of the elementary cone which has P for its apex, and the element dS for its base. The angle, $d\Omega$, may be described as the area which the elementary cone cuts out of the sphere of unit radius described round P as

centre. We have then for the contribution of the element dS to the potential at P:

$$dV = \frac{Ai\,d\Omega}{a}, \quad . \quad . \quad . \quad . \quad . \quad \textbf{(22·2)}$$

and for the potential at P,

$$V = \frac{Ai\Omega}{a}. \quad . \quad . \quad . \quad . \quad \textbf{(22·21)}$$

This last result needs a little further elucidation. First of all, we have fixed the arbitrary constant in V in such a way as to make $V = 0$ at infinitely distant points. This is implied in the use of (**22·02**). Secondly, V would have turned out to be negative at P if this point were on the negative side (as defined by the screw rule) of the sheet. To make the formula apply universally we must therefore use Ω algebraically. That is to say, Ω represents not the numerical value of the solid angle, but the numerical value preceded by $+$ or $-$ according as the point P is on the $+$ or on the $-$ side of the sheet. Lastly, the position and shape of the sheet may be chosen arbitrarily (subject to its having the loop for its boundary); but once chosen it may only be changed in shape or position if certain precautions are taken, and the path of integration of the line integral which represents V *must not pass through the sheet*. It is instructive to consider the potentials at two infinitely near points situated on opposite sides of the sheet. The value of Ω on the positive side of the sheet might be, for example,

$$\Omega_+ = + \tfrac{3}{4} \times 4\pi.$$

Obviously the value of Ω for the other point would then be

$$\Omega_- = - \tfrac{1}{4} \times 4\pi,$$

since it has the opposite sign and in absolute value it represents the remaining or complementary part of the area of the unit sphere. Therefore

$$\Omega_+ - \Omega_- = 4\pi.$$

Consequently

$$V_+ - V_- = \frac{4\pi iA}{a}. \quad . \quad . \quad . \quad \textbf{(22·22)}$$

If therefore the sheet were moved so that a point previously on its negative side might come to be on its positive side, the potential at such a point would increase discontinuously by $4\pi iA/a$. The sheet therefore represents a boundary where V is discontinuous. At all points outside this boundary V is a one-

valued and continuous function of the co-ordinates of the point. If however we abolish the sheet, or, what amounts to the same thing, we assume the privilege of taking the line integral, which defines the potential V, along any path whatever, we find that

$$V = \begin{pmatrix} \text{continuous one-valued} \\ \text{function of } x, y, z \end{pmatrix} + \frac{4\pi i n A}{a}, \quad . \quad . \quad \textbf{(22·23)}$$

where n is any integer, positive or negative. This must be so, since if we take the line integral once round the wire or loop, along which the current, i, is flowing, from the positive side of the sheet to the negative, i.e. in the *positive direction* (Fig. 22) relatively to that of the current, the value of V decreases by $4\pi iA/a$. It increases by the same amount when we take the line integral once round the wire in the *negative direction*. It will be noticed that the terms *positive* and *negative* are used in such a sense that **when we assign the term positive to the direction of the current, the corresponding positive direction of the line integral is related to it in the same way as the direction of rotation of an ordinary screw to that in which it travels.** The formula **(22·23)** is reminiscent of the analogous definition of a logarithm as a line integral.

A very important purpose of the function V is to enable us to calculate the field intensity $\mathbf{H}$ by means of the formula

$$\mathbf{H} = - \text{grad } V.$$

When we apply this formula, the many-valued constant term, $4\pi inA/a$, of **(22·23)** gives simply zero, with the result that, though V is many-valued, $\mathbf{H}$ is one-valued, as of course it should be, since it represents a physical characteristic of the field at the point in question.

§ 22·3.   First Law of Electrodynamics

We conclude from the analysis, given in **§ 22·2**, of the magnetic field of a current flowing round a loop of wire, that

$$\frac{4\pi iA}{a} = \oint (\mathbf{H}, \, d\mathbf{l}), \quad . \quad . \quad . \quad \textbf{(22·3)}$$

where the line integral is taken once round a loop enclosing the wire, and in the positive direction relatively to that of the current. It is easy to see that when we have two or more wires conveying currents, and when the line integral is extended round a closed path embracing all of them, that

$$\frac{4\pi(i_1 + i_2 + i_3 + \ldots)A}{a} = \oint (\mathbf{H}, \, d\mathbf{l}), \quad . \quad \textbf{(22·301)}$$

attention being paid to the sign. It is convenient to express this result in a different way. Imagine a thin sheet or surface, the boundary of which is the closed path round which the integral $\oint$ **(H, dl)** extends. Then $i_1 + i_2 + i_3 + \ldots$ may be written

$$\int\int \textbf{(i, dS)},$$

where **i** means current density, and **dS** is a vectorial element of the surface or sheet. Equation **(22·301)** may therefore be written

$$\boxed{\frac{4\pi A}{a}\int\int \textbf{(i, dS)} = \oint \textbf{(H, dl)}.} \qquad . \quad . \quad . \textbf{(22·31)}$$

We shall adopt all that is implied by Maxwell's displacement hypothesis **(§ 19·2)**, so that the current density **i** may be regarded as the vector sum of one or more parts i_1 and i_2 and i_3, where

$$i_1 = \frac{\partial \textbf{D}}{\partial t}. \quad [\textbf{D} = \text{dielectric displacement}], \quad . \quad \textbf{(22·32)}$$

$$i_2 = \sigma\mathcal{E} [\sigma = \text{conductivity of medium}], \quad . \quad \textbf{(22·33)}$$

$$i_3 = \rho\textbf{v}\begin{bmatrix} \rho = \text{electric density and} \\ \textbf{v} = \text{velocity of transport} \end{bmatrix} . \quad . \quad . \textbf{(22·34)}$$

Equation **(22·31)** has a very fundamental significance. It is sometimes called the **first law of electrodynamics.**

§ 22·4. SECOND LAW OF ELECTRODYNAMICS

Let us turn back to **(22·21)**. It represents the work that *we* must do per unit pole strength to bring up a point north pole from ∞ to the point P (Fig. 22·2) without crossing the barrier of the sheet ; or the work the *field* must do when the pole travels from P to ∞. If instead of being reckoned per unit pole the work is reckoned for a point north pole of strength m, we shall have

$$\text{Work} = \frac{Aim\Omega}{a}. \quad . \quad . \quad . \quad . \textbf{(22·4)}$$

The presence of the pole m at P will give rise to a displacement (in Maxwell's sense) of magnetism through the loop of wire. This must be represented by

$$-\int\int \textbf{(D}_m\textbf{, dS)}, \quad . \quad . \quad . \textbf{(22·41)}$$

the integration being extended over the sheet. The minus sign is necessary because the sense of the displacement is through the sheet in the direction *away* from P, whereas the vector arrow of the element **dS** is directed *towards* P. On the other hand the quantity of magnetism displaced through the loop must be equal to

$$m\Omega/4\pi, \quad . \quad . \quad . \quad . \quad . \quad \textbf{(22·411)}$$

since the total amount m is displaced uniformly in all directions. This amounts to $m/4\pi$ per unit solid angle, and therefore to the quantity expressed by **(22·411)** within the angle Ω. This formula appears to contradict the fundamental principle that the algebraic sum-total of magnetism displaced through any closed surface is zero. The contradiction is however only apparent. The north pole m may be pictured in the way shown in Fig. 22·4. We may regard it as terminating an infinitely thin and long magnet. The quantity m is displaced *inwards* through the part of any closed surface surrounding the pole (such as that represented by the broken line), *within the (infinitely narrow) cross-section of the magnet*, as indicated by the arrows in the figure. While an equal quantity m is displaced *outwards* and uniformly through the remaining part (which is practically the whole) of the closed surface. On equating **(22·41)** and **(22·411)** and substituting in **(22·4)** we get :
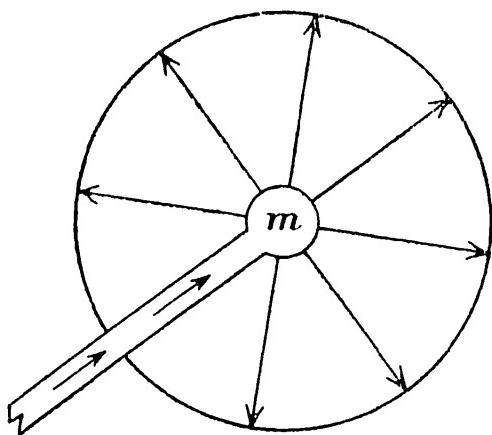
FIG. 22·4

$$\text{Work} = -\frac{A+\pi i}{a}\int\!\!\int (\mathbf{D}_m, \, \mathbf{dS}). \quad . \quad . \quad . \quad \textbf{(22·42)}$$

This then is the amount of work we should have to do to establish the flux $\int\!\!\int (\mathbf{D}_m, \, \mathbf{dS})$ through the loop of wire in which the current i is flowing. In arriving at this result we have assumed i to be kept constant while the flux was being established. Let us now suppose the constant current i to be very small, so that we may regard the energy of that part of the magnetic field due to i as vanishingly small everywhere. The whole energy of the magnetic field may then be supposed to be that of the pole m which contributes the flux $\int\!\!\int (\mathbf{D}_m, \, \mathbf{dS})$. It is obvious that

the energy *of the field* is not altered by the establishment of the flux, while nevertheless work has been done in the process. What form of energy then has been produced, and where? We can only conclude that the work has been done in maintaining or partly maintaining the current in the wire, and, of course, if the circuit consists merely of a piece of wire, the form of energy generated will be heat. We are thus led to the result that the external E.M.F. (due to a cell or other device) necessary to maintain the constant current i will be less during the establishment of the flux $\iint (\mathbf{D}_m, d\mathbf{S})$; that, in fact, while this flux is being produced there is an **induced electromotive force** round the loop of wire and a consequent electric field intensity, at every point of the loop, tending to maintain the current i. Now in accordance with **§ 21·9**, if $\mathcal{E}$ be the intensity of the induced field, the rate at which work is done by it will be

$$i \oint (\mathcal{E} d\mathbf{l}),$$

and by **(22·42)** this must also be equal to

$$- \frac{4\pi i A}{a} \iint \left(\frac{\partial \mathbf{D}_m}{\partial t}, d\mathbf{S} \right).$$

We arrive therefore at the result:

$$\oint (\mathcal{E}, d\mathbf{l}) = - \frac{4\pi A}{a} \iint \left(\frac{\partial \mathbf{D}_m}{\partial t}, d\mathbf{S} \right). \quad . \quad . \quad \textbf{(22·43)}$$

It is often convenient to represent the magnetic flux through a surface by lines (just as we represent the intensity of an electric or magnetic field by lines of force). These are usually called **lines of induction**, and we shall define the number, N, of them, which cut any surface, by

$$N = \iint (\mathbf{D}_m, d\mathbf{S}). \quad . \quad . \quad \textbf{(22·431)}$$

The left-hand member of **(22·43)** represents, as we have seen, the electromotive force induced in the loop of wire, while the right-hand member becomes $- 4\pi A dN/a dt$. Therefore

$$\boxed{\text{Induced Electromotive Force} = - \frac{4\pi A}{a} \frac{dN}{dt}.} \quad \textbf{(22·432)}$$

This is **Faraday's law**; sometimes also termed **Neumann's law**, after Franz Neumann, who first gave mathematical expression to it.

It has become customary—rather unfortunately—to define the number of lines of induction through any surface by

$$N' = \int\int (\mathbf{B}, \mathbf{dS}), \quad . \quad . \quad . \quad (22\cdot433)$$

where

$$\mathbf{B} = \mu\mathbf{H} = 4\pi A \mathbf{D}_m.$$

The quantity **B** (number of lines of induction per unit area) is termed simply the **magnetic induction.** We may consequently write **(22·432)** in the form :

$$\boxed{\text{Induced Electromotive Force} = -\frac{1}{a}\frac{dN'}{dt},} \quad (22\cdot434)$$

and when we adopt electrostatic or electromagnetic units, for which $a = 1$, as we have seen, it simplifies to

$$\boxed{\text{Induced Electromotive Force} = -\frac{dN'}{dt}.} \quad (22\cdot435)$$

The vectors $\mathbf{D}_m$ and **B** become identical when we employ Lorentz-Heaviside units, and in this case, of course, N and N' also become identical

The significance of the sign in **(22·435)** and similar equations must be kept in mind. The left-hand side of this equation represents a quantity directed round a closed loop or curve, while the right-hand side refers to a quantity directed through a sheet or surface which is bounded by the loop. The positive directions for the two quantities are related to one another in the same way as the directions of rotation and of travel of an ordinary right-handed screw. For instance, if the closed curve or loop be in the plane of the paper and the positive direction round it be the clockwise one, the corresponding positive direction for a flux through the paper would be downwards, or away from the reader. The negative sign in **(22·435)** therefore asserts that when the downward flux is actually increasing the induced E.M.F. has a counter-clockwise direction. It has therefore such a direction that it produces, or tends to produce, a magnetic flux in the opposite sense to that which gives rise to it. This is the **law of Lenz.**

The integral

$$\int\int \left(\frac{\partial \mathbf{D}_m}{\partial t}, \mathbf{dS}\right)$$

in (22·43) represents the strength of the magnetic current through the imaginary sheet or surface which we supposed to be bounded by the wire or closed curve round which the line integral $\oint$ ($\mathcal{E}$, dl) is extended. This magnetic current is the exact analogue of an electric displacement current in a dielectric, and no other kind of magnetic current is known. There is no magnetic phenomenon corresponding to conduction or convection currents of electricity. We shall represent the magnetic current density by $\mathbf{i}_m$, so that

$$\mathbf{i}_m = \frac{\partial \mathbf{D}_m}{\partial t}. \qquad \cdot \quad \cdot \quad \cdot \quad \cdot \qquad (22\cdot436)$$

We may now write (22·43) in the form :

$$\boxed{- \frac{4\pi A}{a} \int\!\!\int (\mathbf{i}_m,\, d\mathbf{S}) = \oint(\mathcal{E},\, d\mathbf{l}).} \quad \cdot \quad \cdot \quad (22\cdot44)$$

The line integral is extended round any closed curve, while the surface integral is extended over any surface bounded by the closed curve. We shall term (22·44) the **second law of electro-dynamics.** It should be compared with (22·31).

§ 22·5. Force on a Conductor in which a Current is Flowing

Let AB (Fig. 22·5) represent a wire forming part of a circuit in which a current of strength i is flowing. We shall suppose it to be situated in a magnetic field of external origin. The magnetic displacement at any point of the field may be represented by $\mathbf{D}_m$.

If the wire or circuit be moved so that a short portion of it, dl, sweeps out the small parallelogram shown in the figure, a certain amount of work (positive or negative) will be done. The work done in moving the whole circuit can be obtained by the help of (22·42). Leaving the sign out of account for the moment, this work is equal to the product of $4\pi Ai/a$ and the change in

Fig. 22·5

the value of $\int\!\!\int(\mathbf{D}_m,\, d\mathbf{S})$. Now we may associate a certain part of this work with each element, dl, of the wire as follows :
The change in the value of $\int\!\!\int(\mathbf{D}_m,\, d\mathbf{S})$ is due to the fact that
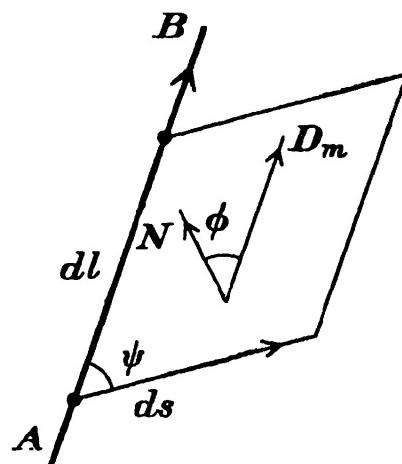
each element, dl, of the wire in suffering a displacement, ds, cuts the magnetic flux. If ψ be the angle between the direction of the current, i, and that of ds, the area swept out by the displacement of dl is

$$dl \, ds \, \sin \, \psi,$$

and if the direction of $\mathbf{D}_m$ makes an angle ϕ with the normal, N, to the plane of the parallelogram, the corresponding contribution to the change of $\iint (\mathbf{D}_m, \mathbf{dS})$ is obviously

$$\mathbf{D}_m \cos \phi \sin \psi \, dl \, ds.$$

Therefore the part of the work done by the displacement of dl is

$$\frac{4\pi A}{a} i \mathbf{D}_m \cos \phi \sin \psi \, dl \, ds.$$

This must be equal to

$$F_s dl \, ds,$$

where F_s is the component of the force (reckoned per unit length of the wire) in the direction ds. On equating the two expressions for the work done on dl, we get

$$F_s = \frac{4\pi A}{a} i \mathbf{D}_m \cos \phi \sin \psi. \quad . \quad . \quad . \quad \textbf{(22·5)}$$

If the displacement, ds, be in the direction of the current, i, the angle ψ, and consequently $\sin \psi$, will be zero, and in this case therefore the component F_s of the force per unit length will also be zero. The force on the wire must therefore be at right angles to its length. Further, if the direction of ds should coincide with that of $\mathbf{D}_m$, the latter will lie in the plane of the parallelogram, and ϕ will be a right angle. Hence the component, F_s, of the force in the direction of $\mathbf{D}_m$ is also zero. We thus arrive at the conclusion that the force on the wire is perpendicular to the plane which contains the directions of the current, i, and the magnetic displacement, $\mathbf{D}_m$. In Fig. 22·51 the current, i, and the displacement, $\mathbf{D}_m$, are represented in the plane of the paper. Let direction of ds be perpendicular to this plane. We need not for the moment decide whether it is directed upwards or downwards. The parallelogram now becomes a rectangle, and
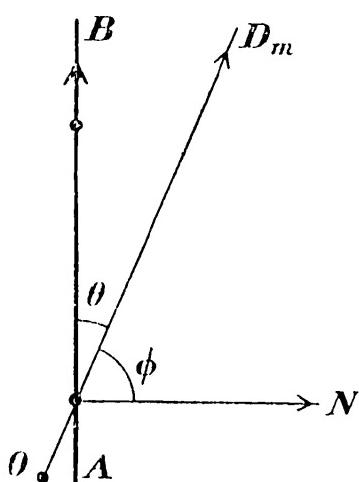
Fig. 22·51

its plane is perpendicular to that of the paper. Since ψ is now a right angle, we find for the force per unit length the expression :

$$\frac{4\pi A}{a} i\mathbf{D}_m \cos \phi,$$

or, if θ be the angle between the directions of the current and of $\mathbf{D}_m$,

$$\frac{4\pi A}{a} i\mathbf{D}_m \sin \theta. \quad . \quad . \quad . \quad . \quad . \quad \textbf{(22·51)}$$

We have thus obtained an expression for the absolute value of the force per unit length ; but we have still to decide whether, in relation to Fig. 22·51, it is directed upwards or downwards. This we can easily do in the following way : Suppose $\mathbf{D}_m$ to be due to a north pole at O. The relations we have already studied (§ 22) between the directions of the current and its magnetic field indicate that the force on this north pole is directed upwards (screw rule). The law of action and reaction therefore requires that the force on the current element must be directed downwards. The expression (22·51) therefore represents a vector product directed downwards. If therefore $\mathbf{F}$ represents the force per unit length of the wire, and if we indicate the vectorial character of the current by $\mathbf{i}$, we find for $\mathbf{F}$:

$$\mathbf{F} = \frac{4\pi A}{a}[\mathbf{i}, \mathbf{D}_m]. \quad . \quad . \quad . \quad . \quad \textbf{(22·52)}$$

Since $\mathbf{D}_m = \mu\mathbf{H}/4\pi A, = \mathbf{B}/4\pi A$ this formula becomes :

$$\mathbf{F} = \frac{1}{a}[\mathbf{i}, \mathbf{B}], \quad . \quad . \quad . \quad . \quad . \quad \textbf{(22·53)}$$

and when we employ electrostatic or electromagnetic units, for which $a = 1$, we have

$$\mathbf{F} = [\mathbf{i}, \mathbf{B}]. \quad . \quad . \quad . \quad . \quad . \quad \textbf{(22·54)}$$

If the medium be air, or empty space, for which μ in E.M. units is unity, or practically unity,

$$\mathbf{F} = [\mathbf{i}, \mathbf{H}] ; \quad . \quad . \quad . \quad . \quad . \quad \textbf{(22·55)}$$

while with *mixed units*, for which $a = c$,

$$\mathbf{F} = \frac{1}{c}[\mathbf{i}, \mathbf{H}]. \quad . \quad . \quad . \quad . \quad . \quad \textbf{(22·56)}$$

It will be well to recall the significance (§ 2·1) of the symbol [ ]. Taking (22·55) for example, [i, H] means a vector the absolute value of which is iH sin θ, where θ is the angle between the directions of i and H. It is also intended to indicate that the direction of this vector is that in which an ordinary screw would travel if it rotated in the sense which would turn i across the angle, θ, towards H.

§ 22·6. FORCE ON A CHARGED PARTICLE IN AN ELECTROMAGNETIC FIELD

We shall now consider a charged particle or small charged body which is moving with a velocity, v, in a magnetic field. We imagine the small body—we take it to be small in order that every part of the magnetic field over which it extends may be considered uniform—
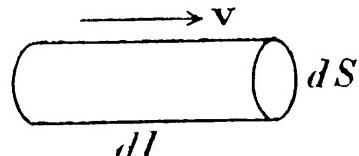
FIG. 22·6

divided into small cylindrical elements of volume, any one of which has a length, dl, in the direction of motion, v, and a cross-sectional area, dS. Let the electric density in the element be ρ. We may regard this element in motion as constituting a current. This current, which is a convection current, is clearly equal to

$$\rho \, dS \, \mathbf{v}.$$

It follows therefore from (22·52) that the force on such an element is

$$\frac{4\pi A}{a} \rho \, dS \, dl \, [\mathbf{v}, \mathbf{D}_m],$$

and since the charge on the element is equal to

$$\rho \, dS \, dl,$$

the force on the element is equal to

$$\text{charge} \times \frac{4\pi A}{a} [\mathbf{v}, \mathbf{D}_m]. \quad . \quad . \quad . \quad (22·6)$$

The factor in (22·6) multiplying the charge on an element is common to all the elements in the body, and consequently the force on the whole body is given by (22·6) if in this formula ' charge ' means the whole charge on the body. We have therefore, if e be the charge on the small body, and F the force :

$$\mathbf{F} = \frac{4\pi A}{a} e[\mathbf{v}, \mathbf{D}_m]. \quad . \quad . \quad . \quad (22·61)$$

This represents, of course, only the force due to the magnetic field. If an electric field of intensity $\mathcal{E}$ be present as well, we have for the complete statement of the force :

$$\boxed{\mathbf{F} = e\left\{\mathcal{E} + \frac{4\pi A}{a}[\mathbf{v}, \mathbf{D}_m]\right\}.}\quad . \quad . \quad . \quad (22\text{·}62)$$

In empty space, and with the use of mixed units, we find, since

$$a = c,$$

and $$\mu_0 = K_0 = 1,$$

$$\mathbf{F} = e\left\{\mathcal{E} + \frac{1}{c}[\mathbf{v}, \mathbf{H}]\right\}. \quad . \quad . \quad . \quad (22\text{·}63)$$

§ 22·7.   Electromagnetic Field Equations

If we apply Stokes's theorem **(3·32)** to the right-hand side of **(22·31)** we get

$$\frac{4\pi A}{a}\int\int (\mathbf{i}, d\mathbf{S}) = \int\int (\mathrm{curl}\ \mathbf{H}, d\mathbf{S}).$$

This equation must hold for any surface, large or small. If therefore the surface be reduced to a mere element $d\mathbf{S}$ we get

$$\frac{4\pi A}{a}(\mathbf{i}, d\mathbf{S}) = (\mathrm{curl}\ \mathbf{H}, d\mathbf{S}),$$

and since no restriction can be placed on the direction of the vector $d\mathbf{S}$, this last equation will be equally true if we turn $d\mathbf{S}$ so that its vector arrow points in the X direction. We have then

$$dS_x = d\mathbf{S},$$
$$dS_y = 0,$$
$$dS_z = 0,$$

and consequently

$$\frac{4\pi A}{a}i_x\, dS_x = [\mathrm{curl}\ \mathbf{H}]_x . dS_x$$

or

$$\frac{4\pi A}{a}i_x = [\mathrm{curl}\ \mathbf{H}]_x.$$

There are, of course, two further equations similarly associated with the Y and Z directions. We shall associate with these three equations the divergence equation **(19·21)**. The appropriateness of this association will presently appear. The four equations may be written:

$$\frac{4\pi A}{a}i_x = \frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z},$$

$$\frac{4\pi A}{a}i_y = \frac{\partial H_x}{\partial z} - \frac{\partial H_z}{\partial x}, \quad . \quad . \quad . \quad . \quad \textbf{(22·7)}$$

$$\frac{4\pi A}{a}i_z = \frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y},$$

$$\rho = \textbf{div D}.$$

Similarly, we get from the second law of electrodynamics **(22·44)** and the associated divergence equation:

$$-\frac{4\pi A}{a}i_{mx} = \frac{\partial \mathcal{E}_z}{\partial y} - \frac{\partial \mathcal{E}_y}{\partial z},$$

$$-\frac{4\pi A}{a}i_{my} = \frac{\partial \mathcal{E}_x}{\partial z} - \frac{\partial \mathcal{E}_z}{\partial x}, \quad . \quad . \quad . \quad \textbf{(22·71)}$$

$$-\frac{4\pi A}{a}i_{mz} = \frac{\partial \mathcal{E}_y}{\partial x} - \frac{\partial \mathcal{E}_x}{\partial y},$$

$$0 = \textbf{div D}_m.$$

We shall direct our attention to a medium with a dielectric constant K; but which is not a perfect insulator. That is to say, it has a conductivity σ which is not zero. Let us further suppose that its permeability μ is a constant. We have then

$$i_x = \frac{K}{4\pi A}\frac{\partial \mathcal{E}_x}{\partial t} + \sigma \mathcal{E}_x,$$

in consequence of **(22·32)** and **(22·33)**, and corresponding expressions for i_y and i_z. For the components of the magnetic current density we find

$$i_{mx} = \frac{\mu}{4\pi A}\frac{\partial H_x}{\partial t}$$

and similar expressions for i_{my} and i_{mz}. Substituting in equations **(22·7)** and **(22·71)** we obtain :

$$\frac{K}{a}\frac{\partial \mathcal{E}_x}{\partial t} + \frac{4\pi A}{a}\sigma \mathcal{E}_x = \frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z},$$

$$\frac{K}{a}\frac{\partial \mathcal{E}_y}{\partial t} + \frac{4\pi A}{a}\sigma \mathcal{E}_y = \frac{\partial H_x}{\partial z} - \frac{\partial H_z}{\partial x}, \qquad . \quad . \quad \textbf{(22·72)}$$

$$\frac{K}{a}\frac{\partial \mathcal{E}_z}{\partial t} + \frac{4\pi A}{a}\sigma \mathcal{E}_z = \frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y},$$

$$\frac{\partial \mathcal{E}_x}{\partial x} + \frac{\partial \mathcal{E}_y}{\partial y} + \frac{\partial \mathcal{E}_z}{\partial z} = \frac{4\pi A}{K}\rho,$$

and

$$-\frac{\mu}{a}\frac{\partial H_x}{\partial t} = \frac{\partial \mathcal{E}_z}{\partial y} - \frac{\partial \mathcal{E}_y}{\partial z},$$

$$-\frac{\mu}{a}\frac{\partial H_y}{\partial t} = \frac{\partial \mathcal{E}_x}{\partial z} - \frac{\partial \mathcal{E}_z}{\partial x}, \qquad . \quad . \quad . \quad \textbf{(22·73)}$$

$$-\frac{\mu}{a}\frac{\partial H_z}{\partial t} = \frac{\partial \mathcal{E}_y}{\partial x} - \frac{\partial \mathcal{E}_x}{\partial y},$$

$$\frac{\partial H_x}{\partial x} + \frac{\partial H_y}{\partial y} + \frac{\partial H_z}{\partial z} = 0.$$

These are the electromagnetic field equations which we owe to Clerk Maxwell.

We shall sometimes use the following notation, or a notation similar to it :

$$H_x \equiv F^{yz}, \qquad \frac{cK}{a}\mathcal{E}_x \equiv F^{lx},$$

$$H_y \equiv F^{zx}, \qquad \frac{cK}{a}\mathcal{E}_y \equiv F^{ly},$$

$$H_z \equiv F^{xy}, \qquad \frac{cK}{a}\mathcal{E}_z \equiv F^{lz},$$

$$s^x \equiv \frac{4\pi A}{a}\sigma \mathcal{E}_x,$$

$$s^y \equiv \frac{4\pi A}{a}\sigma \mathcal{E}_y,$$

$$s^z \equiv \frac{4\pi A}{a}\sigma \mathcal{E}_z,$$

$$s^{l} \equiv \frac{4\pi A}{a}c\rho,$$

and we shall combine with it the convention :

$$F^{xy} = - F^{yx}, \quad F^{xl} = - F^{lx},$$

etc., so that

$$F^{xx} = F^{yy} = F^{zz} = F^{ll} = 0,$$

and we shall use l to represent the product ct. The factor c is a constant with the dimensions of a velocity. We shall regard it, for the present, merely as a device for introducing symmetry into the field equations, and leave till later any inquiry as to whether a physical meaning can be attached to it.

When we introduce this new notation into equations $(22\cdot72)$ they take the rather surprising form :

$$\left. \begin{aligned}
\frac{\partial F^{xx}}{\partial x} + \frac{\partial F^{xy}}{\partial y} + \frac{\partial F^{xz}}{\partial z} + \frac{\partial F^{xl}}{\partial l} &= s^x, \\
\frac{\partial F^{yx}}{\partial x} + \frac{\partial F^{yy}}{\partial y} + \frac{\partial F^{yz}}{\partial z} + \frac{\partial F^{yl}}{\partial l} &= s^y, \\
\frac{\partial F^{zx}}{\partial x} + \frac{\partial F^{zy}}{\partial y} + \frac{\partial F^{zz}}{\partial z} + \frac{\partial F^{zl}}{\partial l} &= s^z, \\
\frac{\partial F^{lx}}{\partial x} + \frac{\partial F^{ly}}{\partial y} + \frac{\partial F^{lz}}{\partial z} + \frac{\partial F^{ll}}{\partial l} &= s^l.
\end{aligned} \right\} \qquad (22\cdot74)$$

We have here again, as in § 10·5, a prevision of the restricted theory of relativity. We learn that we may regard the components of the field vectors as those of a tensor of the second rank in a 4-dimensional continuum, while the current density becomes part of a 4-dimensional vector.

We may deal with the remaining four field equations in a similar way ; but we shall leave this till a later chapter.

If we differentiate equations $(22\cdot74)$ partially, the first with respect to x, the second with respect to y and so on, we find

$$\frac{\partial s^x}{\partial x} + \frac{\partial s^y}{\partial y} + \frac{\partial s^z}{\partial z} + \frac{\partial s^l}{\partial l} = 0, \qquad . \quad . \quad . \quad (22\cdot75)$$

or the 4-dimensional divergence of the vector s is equal to zero. This is easily seen to be equivalent to

$$\mathbf{div}\ (\sigma \mathbf{\mathcal{E}}) + \frac{\partial \rho}{\partial t} = 0, \qquad . \quad . \quad . \quad (22\cdot751)$$

an equation which expresses the conservation of electricity.

CHAPTER VI

INDUCTIVE CIRCUITS

§ 22·8. Self Inductance

WE have seen (22·435) that the induced electromotive force round a closed loop is equal to $- dN'/dt$, when N' means the integral $\iint (\mathbf{B}, \mathbf{dS})$ or $\iint (\mu \mathbf{H}, \mathbf{dS})$ extended over a surface, or imaginary sheet, having the wire for its boundary, and when we adopt units (e.g. E.M. units) for which the numerical constant a is unity. The magnetic induction, $\mathbf{B}$, in the region in which the circuit or loop of wire is situated is commonly represented by lines (lines of induction, § 22·4), the number passing through *any* surface element $\mathbf{dS}$ (in the sense of its vector arrow) being $(\mathbf{B}, \mathbf{dS})$ or $\mathbf{B} \cos \theta \, \mathbf{dS}$. This may also be expressed by saying that the number crossing any small plane surface dS which is perpendicular to the direction of $\mathbf{B}$ is numerically equal to $\mathbf{B}dS$. We can visualize the whole field as mapped out by these lines of induction, and it will be helpful to us to endeavour to describe N' in terms of them. The circuit to which N' refers may be a simple loop of wire in a plane, or it may be coiled in a more or less complicated fashion ; but in either case we think of a surface or imaginary sheet which has the wire for its boundary. If we distinguish the sides of the sheet by the letters α and β, the direction of the vectorial arrow of $\mathbf{dS}$ being from α to β ; and if we further associate the positive sign with all lines of induction cutting the sheet in the sense α to β, and the negative sign with those cutting it in the opposite sense, then the contribution of a single line, of those representing the flux in the field, to the number N' will be the algebraic sum of the number of times it cuts the sheet. It is convenient to call N' the **linked flux** to distinguish it from the flux mentioned above.

Let us now consider the case where the magnetic field is due solely to the current flowing in the loop of wire, and no part

of the field has any other origin. The absolute value of **H** at any point is of course proportional to the current strength, i, and consequently, if we confine our attention to a medium for which μ is constant, the value of **B** at any point is proportional to i. It follows that N', the integrated value of **B** over the sheet, is proportional to i. Therefore

$$N' = Li. \qquad . \quad . \quad . \quad . \quad (22\cdot8)$$

The constant, L, is called the **self inductance** of the circuit or loop of wire. We may of course regard L as localized, and speak of the self inductance of a wire or coil forming part of a circuit, in which case the corresponding N' represents the flux linked with that part of the circuit.

§ 22·9. Mutual Inductance

If N'_{AB} represent the part of the flux linked with a circuit or coil, B, in consequence of a current i_A in another coil, A, we shall have, if the permeability of the medium in which the coils are situated is constant,

$$N'_{AB} = M_{AB}\, i_A, \quad . \quad . \quad . \quad (22\cdot9)$$

where M_{AB} is a constant depending on the positions and shapes of the circuits, and of course a corresponding equation for the flux linked with the coil A in consequence of a current i_B in the coil B; or

$$N'_{BA} = M_{BA}\, i_B. \quad . \quad . \quad . \quad (22\cdot91)$$

In these equations we may choose directions so that both M_{AB} and M_{BA} are positive, and we shall prove in **§ 23·2** that $M_{AB} = M_{BA} = M$. The coefficient M is called the **mutual inductance** of the pair of coils.

§ 23. Some Mathematical Theorems

Let ϕ represent a one-valued function of position, i.e. a function of x, y and z, such for example as electric density or scalar potential, and let ψ mean the function :

$$\psi = \iiint \frac{\phi}{r}\, dx\, dy\, dz, \quad . \quad . \quad . \quad (23)$$

where r is the distance of the point (x, y, z) from a specified point (x_0, y_0, z_0), the integration being extended over all values of x, y and z. The integral is of course assumed to be convergent. What this means may be briefly described as follows : We

imagine the integration to be carried out over the region enclosed between two spheres described about $(x_0,\ y_0,\ z_0)$ as centre, one with a small radius, r_1, and the other with a large radius, r_2. The assumption of convergence means that this integral approaches a finite limiting value as r_1 approaches zero and r_2 approaches infinity. The function ψ is obviously a function of $(x_0,\ y_0,\ z_0)$.

We shall now show that

$$\frac{\partial \psi}{\partial x_0} = \iiint \frac{\frac{\partial \phi}{\partial x}}{r}\ dx\ dy\ dz.\quad .\quad .\quad .\quad (23 \cdot 01)$$

In Fig. 23 two identical volume elements, $dx\ dy\ dz$, are shown; one of them having its centre at $(x,\ y,\ z)$, and the other having its centre at $(x + dx_0,\ y,\ z)$. Obviously the two lines, r, one joining $(x_0,\ y_0,\ z_0)$ to the element at $(x,\ y,\ z)$, and the other joining $(x_0 + dx_0,\ y_0,\ z_0)$ to the element at $(x + dx_0,\ y,\ z)$, are equal and parallel to one another. The first volume element contributes to $\psi\ (x_0,\ y_0,\ z_0)$ the amount

FIG. 23

$$\frac{\phi}{r}\ dx\ dy\ dz\ ;$$

the other contributes to the value of ψ at the point $(x_0 + dx_0,\ y_0,\ z_0)$ the amount

$$\frac{\phi + \frac{\partial \phi}{\partial x}dx_0}{r}\ dx\ dy\ dz,$$

and clearly

$$\psi_{x_0 + dx_0, y_0, z_0} - \psi_{x_0, y_0, z_0} = \iiint \frac{\frac{\partial \phi}{\partial x}dx_0}{r}\ dx\ dy\ dz.$$

Therefore

$$\frac{\partial \psi}{\partial x_0}dx_0 = dx_0 \iiint \frac{\frac{\partial \phi}{\partial x}}{r}\ dx\ dy\ dz,$$

and the formula (23·01) follows in consequence.
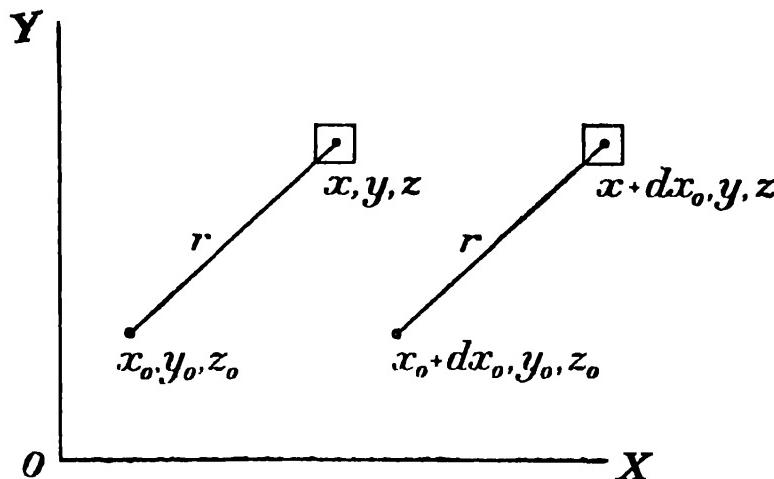
Let us now study the vector

$$\mathbf{s} = \frac{1}{4\pi} \int\int\int \frac{\text{curl } \mathbf{t}}{r} \, dx \, dy \, dz, \quad . \quad . \quad . \quad (23\cdot01)$$

the integration again being extended over all values of x, y and z, and $\mathbf{s}$ meaning the value of the vector at the point (x_0, y_0, z_0) from which the distances r are measured. The theorem (23) gives

$$\text{curl } \mathbf{s} = \frac{1}{4\pi} \int\int\int \frac{\text{curl curl } \mathbf{t}}{r} \, dx \, dy \, dz,$$

and by (2·45) therefore

$$\text{curl } \mathbf{s} = \frac{1}{4\pi} \int\int\int \frac{\text{grad div } \mathbf{t}}{r} \, dx \, dy \, dz - \frac{1}{4\pi} \int\int\int \frac{\nabla^2 \mathbf{t}}{r} \, dx \, dy \, dz.$$

If $\mathbf{t}$ approaches zero sufficiently rapidly with increasing r (see § 3·1), the second integral represents $\mathbf{t}$ at the point (x_0, y_0, z_0), and if further

$$\text{div } \mathbf{t} = 0,$$

we obtain

$$\mathbf{t} = \text{curl } \mathbf{s}. \quad . \quad . \quad . \quad . \quad (23\cdot02)$$

If any other vector, $\mathbf{s}'$, have the property

$$\mathbf{t} = \text{curl } \mathbf{s}',$$

then the two vectors $\mathbf{s}'$ and $\mathbf{s}$ must differ by a vector which is a gradient, i.e.

$$\mathbf{s}' - \mathbf{s} = \text{grad } \phi, \quad . \quad . \quad . \quad (23\cdot03)$$

where ϕ is a scalar quantity. We can demonstrate this in the following way :

$$\text{curl } \mathbf{s}' - \text{curl } \mathbf{s} = 0$$

by hypothesis, and therefore

$$\text{curl } (\mathbf{s}' - \mathbf{s}) = 0,$$

and consequently

$$\int\int\int \frac{\text{curl curl } \boldsymbol{\omega}}{r} \, dx \, dy \, dz = 0,$$

where

$$\boldsymbol{\omega} \text{ means } \mathbf{s}' - \mathbf{s}.$$

The theorem (2·45) gives

$$\iiint \frac{\mathbf{grad}\ \mathbf{div}\ \boldsymbol{\omega}}{r}\ dx\ dy\ dz = \iiint \frac{\nabla^2\boldsymbol{\omega}}{r}\ dx\ dy\ dz,$$

or $\mathbf{grad}\left\{\dfrac{1}{4\pi}\iiint \dfrac{-\mathbf{div}\ \boldsymbol{\omega}}{r}\ dx\ dy\ dz\right\} = \boldsymbol{\omega}.$

This is identical with (23·03) if we write

$$\phi = -\frac{1}{4\pi}\iiint \frac{\mathbf{div}\ (\mathbf{s}' - \mathbf{s})}{r}\ dx\ dy\ dz. \quad . \quad . \quad (23\cdot04)$$

§ 23·1. The Vector Potential

The formula (3·16) is valid for **H** when **H** means the field intensity due to a current circuit of finite dimensions, since in this case it behaves at infinity like $1/r^3$. If, further, the permeability of the medium be constant, i.e. independent of **H**, div **H** = 0, and we may make use of (23·02) and write

$$\mathbf{H} = \mathbf{curl}\ \mathbf{A}. \quad . \quad . \quad . \quad . \quad (23\cdot1)$$

If we adopt for **A** the vector

$$\mathbf{A} = \frac{1}{4\pi}\iiint \frac{\mathbf{curl}\ \mathbf{H}}{r}\ dx\ dy\ dz \quad . \quad . \quad . \quad (23\cdot2)$$

(see (23·01)), it will have the property

$$\mathbf{div}\ \mathbf{A} = 0,$$

since, by (2·42), the divergence of a curl is identically zero.

Turning to equations (22·7) and substituting **curl** **A** for **H**, we find at once

$$\frac{4\pi A}{a} i_x = -\nabla^2 A_x,$$

$$\frac{4\pi A}{a} i_y = -\nabla^2 A_y, \quad . \quad . \quad . \quad . \quad (23\cdot11)$$

$$\frac{4\pi A}{a} i_y = -\nabla^2 A_z,$$

since **div** **A** = 0.

If now the current density **i**, and consequently its components i_x, i_y and i_z, are functions of position and independent of the time, these equations are mathematically identical with

Poisson's equation (§ 18·7). The solution of the first one, for example, may be written :

$$A_x(x_0,\ y_0,\ z_0) = \frac{A}{a} \int\int\int \frac{i_x}{r}\ dx\ dy\ dz, \qquad . \quad . \quad (23\cdot12)$$

the integration extending over all the region where i_x differs from zero.

We may apply this result to the special case of a single loop of wire carrying a steady current. Instead of the volume element $dx\ dy\ dz$ we shall employ an element like that shown in Fig. 22·6, which we may now regard as representing a piece of the wire through which the current is flowing, dl in length and dS in cross-sectional area. Therefore the contribution of any element dl to the value of the component A_s of the vector $\mathbf{A}$ at some point $x_0,\ y_0,\ z_0$ is

$$\frac{Ai\ \cos\ \phi\, dl}{ar},$$

where $i(=\mathbf{i}\ dS)$ is the total current in the wire, and must be sharply distinguished from the current density $\mathbf{i}$; and ϕ is the angle between the direction of the current, i, and the direction s. We have consequently for A_s the expression :

$$A_s = \frac{A}{a} \oint \frac{i\ \cos\ \phi\ dl}{r}, \qquad . \quad . \quad . \quad . \quad (23\cdot13)$$

in which, as the symbol $\oint$ indicates, the integration is to be extended round the current loop.

§ 23·2. Neumann's Theorem

We now consider two circuits—we may suppose them to be two loops of wire—round one of which the current i is flowing, and we inquire about the flux linked through the other. That is to say, if as usual we suppose each loop to be the boundary of a surface or sheet, we wish to find an expression for the algebraic number of lines of induction cutting the second sheet. We shall confine our attention to the case where the permeability of the medium in which the circuits are situated is constant (i.e. independent of $\mathbf{H}$), and distinguish the two circuits by the numbers 1 and 2. Then the linked flux will be :

$$Mi\ = \text{constant} \times \int\int (\mathbf{H},\ \mathbf{dS}),$$

the constant being $\mu/4\pi A$. $\mathbf{H}$ is the magnetic field intensity

at points on sheet 2 due to the current, i, in the circuit 1, and $\mathbf{dS}$ is an element of the surface 2. By (23·1) this may be written :

$$Mi = \frac{\mu}{4\pi A} \iint (\mathbf{curl\ A, dS}),$$

and therefore (Stokes's theorem)

$$Mi = \frac{\mu}{4\pi A} \oint (\mathbf{A, dl}).$$

This is equivalent to

$$Mi = \frac{\mu}{4\pi A} \oint A_2 dl_2, \qquad . \quad . \quad . \quad . \quad \textbf{(23·2)}$$

where A_2 is the component of $\mathbf{A}$ in the direction dl_2. By (23·13)

$$A_2 = \frac{A}{a} \oint \frac{i \cos \phi \, dl_1}{r}, \qquad . \quad . \quad . \quad . \quad \textbf{(23·201)}$$

where r is the distance between the elements dl_1 and dl_2, and ϕ is the angle between their directions. On substituting in (23·2) we get

$$Mi = \frac{\mu}{4\pi a} \oint \oint \frac{i \cos \phi \, dl_1 dl_2}{r}.$$

Whence

$$M = \frac{\mu}{4\pi a} \oint \oint \frac{\cos \phi \, dl_1 dl_2}{r}. \quad . \quad . \quad . \quad \textbf{(23·21)}$$

The current, i, is flowing round circuit 1 in the direction dl_1. If instead we have a current, i, flowing round circuit 2 in the direction of dl_2, it is quite obvious that we get precisely the same expression, (23·21), for the consequent flux linked through circuit 1, which proves the statement $M_{AB} = M_{BA}$ (22·9 and 22·91). This is known as **Neumann's Theorem.**

§ 23·3. MUTUAL INDUCTANCE OF TWO COAXIAL CIRCLES

We shall apply Neumann's formula (23·21) to the special case where the two circuits are circles with a common axis. Having adopted a direction along the common axis (a in Fig. 23·3A) as the positive direction, we shall carry out the two integrations of Neumann's formula in the positive sense round both circles. Let L be the axial distance between the two

circles, R_1 and R_2 their respective radii, and ϕ the angle shown in Fig. 23·3B, in which the two circles are shown projected on
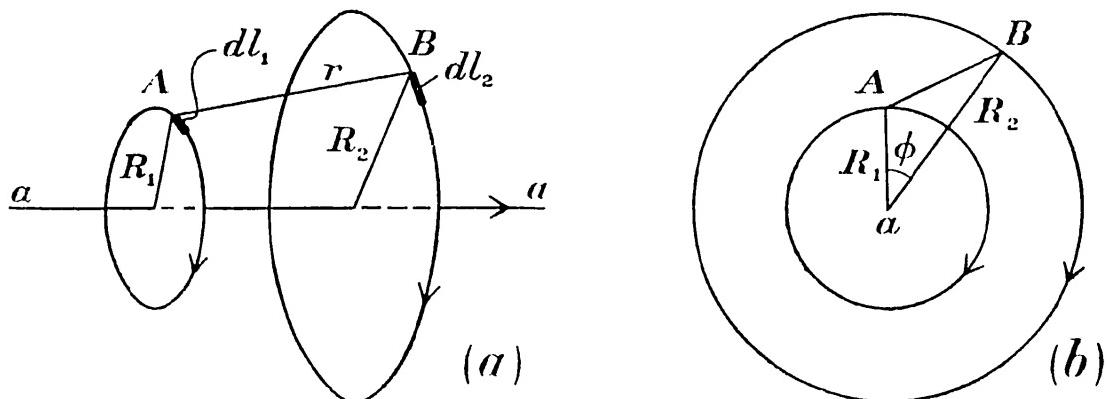
FIG. 23·3

a plane perpendicular to the axis. The square of the projection of the line r, joining the points A and B, is

$$R_1{}^2 + R_2{}^2 - 2R_1R_2 \cos \phi,$$

and consequently

$$r = \{L^2 + R_1{}^2 + R_2{}^2 - 2R_1R_2 \cos \phi\}^{1/2}. \qquad . \quad (23·3)$$

The tangential component, A_2, of the vector potential at any point, B, on the circle, 2, due to the unit current in circle, 1, is (23·201), with units for which $A = a = 1$,

$$A_2 = R_1 \oint_1 \frac{\cos \phi \, d\phi}{\{L^2 + R_1{}^2 + R_2{}^2 - 2R_1R_2 \cos \phi\}^{1/2}},$$

since $$dl_1 = R_1 d\phi.$$

It is easy to see that A_2 has the same value at all points on circuit 2, so that the further integration (23·21) with respect to l_2 is simply equivalent to multiplying by $2\pi R_2$. The flux of *magnetic force* through 2 is therefore

$$2\pi R_1 R_2 \int_0^{2\pi} \frac{\cos \phi \, d\phi}{\{L^2 + R_1{}^2 + R_2{}^2 - 2R_1R_2 \cos \phi\}^{1/2}}. \qquad . \quad (23·31)$$

If the permeability of the medium be unity, this will be identical with the mutual inductance of the circles, as defined in § 22·9.

To evaluate (23·31) we introduce a new variable, θ, defined by

$$\phi = \theta + \pi,$$

and on substituting we get

$$M = 2\pi R_1 R_2 \int_{-\pi}^{+\pi} \frac{- \cos \theta . d\theta}{\{L^2 + R_1{}^2 + R_2{}^2 + 2R_1R_2 \cos \theta\}^{1/2}}.$$

On making the further substitution

$$2\varepsilon = \theta,$$

$$M = 4\pi R_1 R_2 \int_{-\pi/2}^{+\pi/2} \frac{(2\sin^2 \varepsilon - 1)d\varepsilon}{\{L^2 + (R_1 + R_2)^2 - 4R_1R_2 \sin^2 \varepsilon\}^{1/2}},$$

or $\quad M = \dfrac{4\pi R_1 R_2}{\{L^2 + (R_1 + R_2)^2\}^{1/2}} \displaystyle\int_{-\pi/2}^{+\pi/2} \dfrac{(2\sin^2 \varepsilon - 1)d\varepsilon}{\{1 - \kappa^2 \sin^2 \varepsilon\}^{1/2}},$

in which

$$\kappa^2 \equiv \frac{4R_1R_2}{\{L^2 + (R_1 + R_2)^2\}}.$$

This may be written

$$M = 4\pi\sqrt{R_1R_2}\,|\,\kappa \int_0^{\pi/2} \frac{(2\sin^2 \varepsilon - 1)d\varepsilon}{\sqrt{1 - \kappa^2 \sin^2 \varepsilon}\,|},$$

which is easily seen to be identical with

$$M = 4\pi\sqrt{R_1R_2}\,|\left\{\left(\frac{2}{\kappa} - \kappa\right)F(\kappa) - \frac{2}{\kappa}E(\kappa)\right\}, \quad . \quad \textbf{(23·32)}$$

where $F(\kappa)$ and $E(\kappa)$ are the elliptic integrals of **§ 7**:

$$F(\kappa) \equiv \int_0^{\pi/2} \frac{d\varepsilon}{\sqrt{1 - \kappa^2 \sin^2 \varepsilon}\,|},$$

$$E(\kappa) \equiv \int_0^{\pi/2} \sqrt{1 - \kappa^2 \sin^2 \varepsilon}\,|\,d\varepsilon.$$

The modulus κ is, of course, less than unity, and if we write

$$\kappa = \sin \gamma,$$

we find $\quad\quad \cos^2 \gamma = \dfrac{L^2 + (R_1 - R_2)^2}{L^2 + (R_1 + R_2)^2},$

or $\quad\quad \cos \gamma = r'/r'', \quad . \quad . \quad . \quad . \quad . \quad \textbf{(23·321)}$

where r' means the shortest distance separating two such points as A and B (Fig. 23·3), and r'' the longest possible distance. The logarithm of expression,

$$\left(\frac{2}{\kappa} - \kappa\right)F(\kappa) - \frac{2}{\kappa}E(\kappa), \quad . \quad . \quad . \quad \textbf{(23·33)}$$

has been tabulated by Clerk Maxwell for values of γ ranging from 60° to 90° (*Treatise*, Vol. II, second edition, p. 317).

If the square root expressions $(1 - \kappa^2 \sin^2 \varepsilon)^{-1/2}$ and $(1 - \kappa^2 \sin^2 \varepsilon)^{\frac{1}{2}}$ in the integrals F and E respectively be expanded by the binomial theorem, and the integrations then carried out, it will be found that the resulting expression for (23·33) approaches the limit

$$\frac{\pi\kappa^3}{16}, \qquad \cdots \qquad \cdots \qquad (23\cdot34)$$

when κ approaches zero. We may verify this result by applying it to two very simple cases : firstly, that in which circle 2 is coplanar with circle 1 ($L = 0$), and of infinitesimal radius (R_2 very small). Elementary considerations make

$$M = 2\pi^2 R_2{}^2/R_1, \qquad \cdots \qquad (23\cdot341)$$

while (23·32) in association with (23·34) yields

$$M = 4\pi \sqrt{R_1 R_2} \left| \frac{\pi\kappa^3}{16} \right. \qquad \cdots \qquad (23\cdot342)$$

In this case $\qquad \kappa^2 = \dfrac{4R_1 R_2}{(R_1 + R_2)^2} = 4\dfrac{R_2}{R_1},$

and therefore $\qquad \kappa^3 = 8\sqrt{\left.\dfrac{R_2{}^3}{R_1{}^3}\right|}.$

On substituting this in (23·342) we obtain the result (23·341) as we should expect. The other simple illustration is that for which the circles are separated by a distance very great compared with either radius (L very great). Elementary methods yield in this case :

$$M = \frac{2\pi^2 R_1{}^2 R_2{}^2}{L^3},$$

while the same result will be seen to emerge from (23·32) in association with (23·34).

§ 23·4.   Energy in the Magnetic Fields of Currents

We shall now investigate the energy in the magnetic field of a current of strength, i, flowing round a loop of wire abc (Fig. 23·4). We may think of the part of the loop represented by the broken line as below the plane of the paper. Imagine

a sheet with *abc* as its boundary, and suppose the magnetic lines of force constructed which cut the sheet at points on the boundary of a surface element, dS_0, of the sheet. These lines will mark out a **tube of force**, *def*. We shall begin by writing down an expression for the magnetic energy in an element, O, of the tube. In accordance with **§ 21·4**, this will amount to

$$\tfrac{1}{2}\mathbf{D}_m\mathbf{H}\, dS\, dl,$$

where dl and dS are respectively the length and cross-sectional area of the element, or

FIG. 23·4

$$\tfrac{1}{2}\mathbf{D}_m dS\, (\mathbf{H},\, \mathbf{dl}), \quad . \quad . \quad . \quad . \quad (23\text{·}4)$$

in which **dl** represents the linear element as a vector having the same direction as **H**. The total energy in the tube will be

$$\tfrac{1}{2}\mathbf{D}_m dS \oint (\mathbf{H},\, \mathbf{dl}), \quad . \quad . \quad . \quad (23\text{·}41)$$

since $\mathbf{D}_m dS$ has (**§ 21**) the same value at all cross-sections of the tube. But by (**22·3**) this is equal to

$$\frac{4\pi A i}{a} \times \tfrac{1}{2}\mathbf{D}_m dS,$$

or

$$\frac{4\pi A i}{a} \times \tfrac{1}{2}(\mathbf{D}_{mo},\, \mathbf{dS}_0),$$

where $\mathbf{dS}_0$ is the surface element cut by the tube out of the sheet which we have supposed to be stretched across *abc*, and $\mathbf{D}_{mo}$ is the magnetic displacement at points on the sheet. When we add up the energies of all such tubes associated with the circuit we get

$$\tfrac{1}{2}\frac{i}{a} \times 4\pi A \int\int (\mathbf{D}_{mo},\, \mathbf{dS}_0),$$

or

$$U = \tfrac{1}{2}\frac{i}{a} \times 4\pi A N,$$

or

$$U = \tfrac{1}{2}\frac{i}{a}N',$$

$$\left.\begin{array}{c}\\[1.2em]\end{array}\right\} \quad . \quad . \quad . \quad . \quad (23\text{·}42)$$
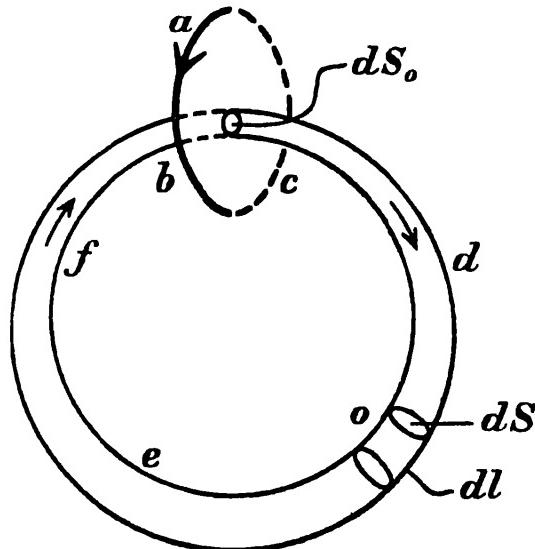
where U is the energy and N and N' have the meanings already described. In accordance with the definition **(22·8)** this is equivalent to

$$U = \tfrac{1}{2}\frac{Li^2}{a}, \qquad \ldots \quad \textbf{(23·43)}$$

or, if we employ units for which $a = 1$,

$$U = \tfrac{1}{2}Li^2. \qquad \ldots \quad \textbf{(23·431)}$$

§ 23·5. Field Energy Associated with Two Circuits

Let the two circuits be represented by closed loops 1 and 2 in Fig. 23·5, the broken parts of the lines representing the parts of the loops below the plane of the paper. Let us further suppose currents i_1 and i_2 flowing round the loops 1 and 2 respectively in the directions indicated in the figure. Some of the lines (or tubes) of induction will travel so as to go round one loop only (a in Fig. 23·5); others, like b, will embrace both loops. Let us imagine a surface, cutting the plane of the paper in the line AB, and dividing the whole space where the magnetic field is appreciable into two parts, one containing the circuit 1 and the other containing the circuit 2. In dealing with such tubes of induction

Fig. 23·5

as a, which envelop the current i_1 only, we can proceed exactly as in **§ 23·4.** In dealing with tubes like b which envelop both currents we shall divide the line integral $\oint$ **(H, dl)** which appears in the expression **(23·41)** into two parts—one part extending along the tube from y to x, points where it is cut by the surface AB, and the remaining part along the tube from x to y, on the other side of AB, and we shall not introduce any error if we add to the first part of the integral any quantity, ε, provided we subtract ε from the second part. We shall take for ε the integral $\int$ **(H, dl)** along a line in the surface AB from x to y. The complete integral $\oint$ **(H, dl)** round the tube may thus be expressed as the sum of two integrals, one en-

circling i_1 only, the other going round i_2 only. Remembering this we see that we may write for the whole energy

$$U = \frac{1}{2a}\{i_1 N_1' + i_2 N_2'\}. \quad . \quad . \quad . \quad . \quad \textbf{(23·5)}$$

where N_1' is the number of lines of induction linked through circuit 1, while N_2' is the number linked through circuit 2. Now if the permeability of the medium in which the circuits are situated be constant (independent of **H**),

$$\left.\begin{array}{l} N_1' = L_1 i_1 + M i_2, \\ N_2' = M i_1 + L_2 i_2, \end{array}\right\} \quad . \quad . \quad . \quad \textbf{(23·51)}$$

where L_1 is the self-inductance of circuit 1, M is the mutual inductance of the circuits and L_2 the self-inductance of circuit 2. On substituting in **(23·5)** we get

$$U = \frac{1}{2a}\{L_1 i_1^2 + 2M i_1 i_2 + L_2 i_2^2\}. \quad . \quad . \quad \textbf{(23·52)}$$

§ 23·6. Alternating Currents

Let us consider a circuit like that illustrated in Fig. 23·6, in which we have resistance, self-inductance and capacity in series, and a cell or other source of current with an E.M.F. equal to E in the sense of the arrow, which in the first instance we shall assume to be constant. Let I be the current at any instant in the direction of the arrows. We have then an E.M.F. in the cell equal to E, and a consequent E.M.F.,

Fig. 23·6

due to the varying magnetic field, equal to $- L\dfrac{dI}{dt}$ in suitable units (**(22·435)** and **(22·8)**), and if we represent the potentials of the condenser plates by O and V, as indicated in the figure, the net work done in overcoming the resistance of the conducting parts of the circuit will be
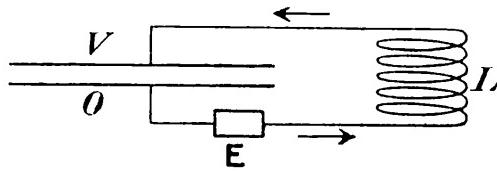
$$\text{E} - L\frac{dI}{dt} - V \quad . \quad . \quad . \quad . \quad \textbf{(23·6)}$$

on each unit quantity of electricity flowing round the circuit. In accordance with the definition of resistance, this expression must be equated to RI. Thus we have

$$\text{E} - L\frac{dI}{dt} - V = RI. \quad . \quad . \quad . \quad \textbf{(23·61)}$$

If Q be the quantity of electricity which has reached the condenser at the instant t,

$$V = Q/C,$$

where C is the capacity of the condenser, and

$$I = dQ/dt \; ;$$

consequently equation **(23·61)** becomes

$$\mathrm{E} = L\frac{d^2Q}{dt^2} + R\frac{dQ}{dt} + \frac{1}{C}Q,$$

or

$$\frac{d^2Q}{dt^2} + \frac{R}{L}\frac{dQ}{dt} + \frac{1}{LC}Q = \frac{E}{L}. \quad . \quad . \quad . \quad \textbf{(23·62)}$$

This equation gives us information as to the dependence of the condenser charge on the time.

If we differentiate **(23·62)** with respect to the time, remembering that E is constant, we easily get

$$\frac{d^2I}{dt^2} + \frac{R}{L}\frac{dI}{dt} + \frac{1}{LC}I = 0. \quad . \quad . \quad . \quad \textbf{(23·63)}$$

This equation enables us to express the current I as a function of the time.

These equations, **(23·62)** and **(23·63)**, are the familiar equations of damped simple harmonic motion. Both of them can be written in the form

$$\frac{d^2x}{dt^2} + 2a\frac{dx}{dt} + b^2x = 0, \quad . \quad . \quad . \quad \textbf{(23·64)}$$

in which a and b^2 are real and positive constants. The general solution of **(23·64)** is

$$x = Ae^{\lambda_1 t} + Be^{\lambda_2 t}, \quad . \quad . \quad . \quad . \quad \textbf{(23·641)}$$

where
$$\lambda_1 \equiv -a + \sqrt{a^2 - b^2}\,|\,,$$
$$\lambda_2 \equiv -a - \sqrt{a^2 - b^2}\,|\,,$$

and A and B are constants of integration. The expression **(23·641)** fails to be a general solution in the special case where $b^2 = a^2$. It may readily be verified that, in this case, the general solution is

$$x = (A + Bt)e^{-at}, \quad . \quad . \quad . \quad \textbf{(23·642)}$$

where again A and B are constants of integration.

From the point of view associated with our electrical problem

the most interesting case is that in which $b^2 > a^2$, and the λ's consequently complex numbers. In this case it is convenient to write

$$\lambda_1 \equiv -a + i\sqrt{b^2 - a^2}\,|,$$
$$\lambda_2 \equiv -a - i\sqrt{b^2 - a^2}\,|,$$

where i means $\sqrt{-1}\,|$. We shall use the letter ω_0 in the sense

$$\omega_0 \equiv \sqrt{b^2 - a^2}\,|,$$

so that the general solution (23·641) becomes

$$x = e^{-at}\{Ae^{i\omega_0 t} + Be^{-i\omega_0 t}\},$$

or

$$x = e^{-at}\{A' \cos \omega_0 t + B' \sin \omega_0 t\}, \quad . \quad . \quad \textbf{(23·65)}$$

where A' and B' are constants of integration. If we multiply this expression outside the brackets { } by $K(\equiv \sqrt{A'^2 + B'^2}\,|\,)$ and divide by K within the brackets, we get

$$x = Ke^{-at} \cos (\omega_0 t - \eta), \quad . \quad . \quad . \quad \textbf{(23·651)}$$

where $\tan \eta = B'/A'.$

In this expression we may regard K and η as new arbitrary constants. When the **damping factor,** a, is zero, x is a strictly periodic (simple harmonic) function of t, its period being

$$\tau_0 = 2\pi/\omega_0. \quad . \quad . \quad . \quad . \quad \textbf{(23·66)}$$

Even when a is not zero, and x consequently not a *strictly* periodic function of t, it is usual to speak of $\tau(\equiv 2\pi/\omega_0)$ as the **period of oscillation** of x. The quantity, Ke^{-at}, is the **amplitude** of x. It will easily be seen that, so long as a is not less than b, x approaches zero the more rapidly the *smaller* a is (23·641). On the other hand, so long as a is not greater than b, the amplitude of x, namely Ke^{-at}, diminishes the more rapidly the *greater* is the value of a. In fact the disturbance, of which x is a measure, disappears most rapidly when the damping factor, a, has its **critical value,** namely $a = b$.

Equation (23·62) becomes identical with (23·64) when we substitute

$$x \quad \text{for} \quad Q - EC,$$
$$a \quad \text{for} \quad R/2L,$$

and

$$b^2 \quad \text{for} \quad 1/LC.$$

Consequently

$$Q - EC = e^{-Rt/2L} K \cos \left\{ \sqrt{\frac{1}{LC} - \frac{R^2}{4L^2}}\,\Bigg|\, t - \eta \right\}. \quad . \quad \textbf{(23·67)}$$

The constants K and η, or A' and B', are determined by the initial conditions. Let us take. for example, the case where the impressed E.M.F. is zero, i.e. $E = 0$, and as initial conditions :

$$Q = Q_0$$

and
$$\left(\frac{dQ}{dt}\right)_{t=0} = 0.$$

This is the state of affairs at a moment when the plates of a charged condenser are connected through a conducting wire possessing self-inductance. Taking the form of solution **(23·65)**,

$$x = Q - EC = Q = Q_0$$

at the instant $t = 0$. Therefore

$$Q_0 = A'. \quad . \quad . \quad . \quad . \quad \textbf{(23·671)}$$

Differentiating **(23·65)** we get

$$\frac{dx}{dt} = - ax + e^{-at}\{- \omega_0 A' \sin \omega_0 t + \omega_0 B' \cos \omega_0 t\},$$

and
$$\left(\frac{dx}{dt}\right)_{t=0} = \left(\frac{dQ}{dt}\right)_{t=0} = 0 ;$$

therefore
$$0 = - aA' + \omega_0 B'. \quad . \quad . \quad . \quad \textbf{(23·672)}$$

Consequently
$$A' = Q_0,$$
$$B' = RQ_0/2L\sqrt{1/LC - R^2/4L^2}\,|.$$

§ 23·7. Forced Oscillations—Resonance

If the impressed E.M.F. in **(23·61)**, instead of being constant, have the form :

$$E = E_0 \cos \omega t,$$
or
$$E = E_0 \sin \omega t,$$

where E_0 and ω are constants, the former being the amplitude, the latter the **angular frequency**, of the impressed E.M.F., we get on substituting :

$$\text{(A)} \qquad E_0 \cos \omega t - L\frac{dI_1}{dt} - V_1 = RI_1, \left.\begin{matrix} \\ \\ \\ \\ \end{matrix}\right\}$$

$$\text{(B)} \qquad E_0 \sin \omega t - L\frac{dI_2}{dt} - V_2 = RI_2, \left.\begin{matrix} \\ \end{matrix}\right\} \quad . \quad . \quad \textbf{(23·7)}$$

I_1 and I_2 representing the current strengths in the two cases

respectively. We shall multiply (23·7 (B)) by $i = \sqrt{-1}\,|$, and add (A) and (B) together. We thus get

$$E_0 e^{i\omega t} - L\frac{d}{dt}(I_1 + iI_2) - (V_1 + iV_2) = R(I_1 + iI_2),$$

or, if we write

$$I \equiv I_1 + iI_2,$$
$$V \equiv V_1 + iV_2,$$
$$Q \equiv Q_1 + iQ_2,$$

and differentiate :

$$i\omega E_0 e^{i\omega t} - L\frac{d^2 I}{dt^2} - \frac{1}{C}I = R\frac{dI}{dt}.$$

We shall write this in the form :

$$\frac{d^2 I}{dt^2} + 2a\frac{dI}{dt} + b^2 I = \frac{i\omega E_0}{L}e^{i\omega t}, \qquad . \quad . \quad (23\cdot71)$$

in which, it will be observed, a and b have the same meanings as in § 23·6.

This equation has the particular solution :

$$I = Me^{i\omega t}, \qquad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (23\cdot72)$$

provided

$$M \equiv E_0/\{R + i(\omega L - 1/\omega C)\}, \quad . \quad . \quad (23\cdot721)$$

or

$$M \equiv E_0 e^{-i\phi}/\{R^2 + (\omega L - 1/\omega C)^2\}^{1/2},$$

where

$$\tan\phi \equiv (\omega L - 1/\omega C)/R. . \quad . \quad . \quad . \quad . \quad (23\cdot722)$$

The general solution of the equation is :

$$I = Ae^{\lambda_1 t} + Be^{\lambda_2 t} + Me^{i\omega t}, \quad . \quad . \quad . \quad (23\cdot73)$$

where A and B are the constants of integration, and λ_1 and λ_2 have the meanings described in § 23·6. It will be noticed that, after the lapse of a sufficiently long time, the **complementary** part of this solution, namely

$$Ae^{\lambda_1 t} + Be^{\lambda_2 t},$$

becomes negligibly small by comparison with the **particular solution (23·72)**, so that usually I is described by the latter with sufficient accuracy. In this case then we have

$$I = \frac{E_0}{\left\{R^2 + \left(\omega L - \frac{1}{\omega C}\right)^2\right\}^{1/2}}e^{i(\omega t - \phi)}. \quad . \quad . \quad (23\cdot73)$$

The real part of this is the solution of **(23·7 (A))** and the purely imaginary part is the solution of **(23·7 (B))**. Thus

$$\left.\begin{array}{l} I_1 = I_0 \cos (\omega t - \phi), \\ I_2 = I_0 \sin (\omega t - \phi), \end{array}\right\} \qquad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \textbf{(23·74)}$$

where $\qquad I_0 = \mathrm{E}_0/\{R^2 + (\omega L - 1/\omega C)^2\}^{1/2}$

is the amplitude of the current. The quantity, $\omega L - 1/\omega C$, is called the **reactance** of the circuit and

$$\sqrt{R^2 + (\omega L - 1/\omega C)^2} \,|$$
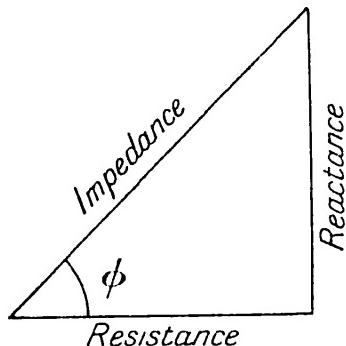
is its **impedance.** The angle, ϕ, represents the amount by which the current lags in phase behind the impressed E.M.F. The relationship between these quantities is illustrated in Fig. 23·7.

Fig. 23·7

If we regard the current amplitude, I_0, as a function of the angular frequency, ω, we observe that it is a maximum when ω has the value $\sqrt{1/LC}\,|$, the reactance, and consequently tan ϕ, then being zero. This special case is called **resonance.** It occurs when the period of the impressed E.M.F. is

$$\tau = 2\pi\sqrt{1/LC}\,| \,.$$

If at the same time the resistance, R, happens to be relatively small, the period

$$\tau_0 (= 2\pi\sqrt{1/LC - R^2/4L^2}\,|)$$

of the **natural oscillations** in the circuit is practically equal to τ, the period of the **forced oscillations.** Resonance occurs then, provided the coefficient, $2a$, in the so-called **dissipative** term of **(23·64)** is negligibly small, when the period of the impressed E.M.F. is equal to, or nearly equal to, the natural period of oscillation of the circuit.

§ 23·8   Activity of the Impressed E.M.F.—Root Mean Square Values

The work done by the impressed E.M.F. during the whole period, $\tau = 2\pi/\omega$, is expressed by

$$\int_0^{2\pi/\omega} \mathrm{E}I dt = \mathrm{E}_0 I_0 \int_0^{2\pi/\omega} \cos \omega t \cos (\omega t - \phi) \, dt, \qquad \cdot \quad \textbf{(23·8)}$$

in consequence of **(23·7)** and **(23·74)**. The integral is equal to

$$\int_{0}^{2\pi/\omega} \cos \omega t \cos (\omega t - \phi)\, dt = \frac{\pi}{\omega} \cos \phi. \quad . \quad . \quad \textbf{(23·81)}$$

Consequently the work done by the impressed E.M.F. during a whole period, $\tau = 2\pi/w$, is

$$E_0 I_0 \frac{\pi}{\omega} \cos \phi,$$

or

$$\frac{E_0 I_0 \tau \cos \phi}{2}. \quad . \quad . \quad . \quad . \quad \textbf{(23·82)}$$

The average rate, therefore, of supply of energy to such a circuit is

$$\frac{E_0 I_0}{2} \cos \phi. \quad . \quad . \quad . \quad . \quad \textbf{(23·83)}$$

The average value of E_0^2 is

$$\frac{E_0^2}{\tau} \int_{0}^{2\pi/\omega} \cos^2 \omega t\, dt,$$

and since, by **(23·81)**, the integral is equal to π/ω, or to $\tau/2$, we have

$$\overline{E^2} = E_0^2/2. \quad . \quad . \quad . \quad . \quad \textbf{(23·84)}$$

Similarly,

$$\overline{I^2} = I_0^2/2. \quad . \quad . \quad . \quad \textbf{(23·841)}$$

If we write

$$\left.\begin{aligned} \sqrt{\overline{E^2}} &= E_R, \\ \sqrt{\overline{I^2}} &= I_R, \end{aligned}\right\} \quad . \quad . \quad . \quad \textbf{(23·842)}$$

we find that **(23·83)** becomes

$$E_R I_R \cos \phi. \quad . \quad . \quad . \quad . \quad \textbf{(23·85)}$$

The factor $\cos \phi$ is called the **power factor**, while E_R and I_R are termed the **root mean square** values of the E.M.F. and current, respectively. We may therefore say that **the rate of supply of energy to the circuit is equal to the product of the power factor and the root mean square values of the E.M.F. and the current.**

It is a consequence of **(23·85)** that the rate of supply of energy approximates to zero as ϕ approaches $+$ or $-\pi/2$. This happens when the resistance is very small compared with the reactance. In the circuit we are studying, the energy supplied is converted into heat in conducting part of it. The value of I_R can, for given E_R, be kept down to any desired limit by arranging that the impedance in the circuit is big enough ; and this can be done by making the reactance large—by a large value of L for example—while the resistance R may be quite small. That is to say, the current may be kept below some desired limit by a means which makes the power factor small and thus avoids the waste of energy in the form of heat. This is the principle of the **choking coil**, which is simply a coil of variable self-inductance and low resistance. It should be observed that the power factor becomes equal to unity when resonance sets in, since the reactance is then zero.

The consequences of making the capacity, which is in series in a circuit, very large are (a) that the amplitude of V, the potential difference between the condenser plates, will become very small, since V is always equal to Q/C, and (b) that the term $1/\omega C$ in the reactance likewise becomes small. Consequently if the condenser be removed, i.e. if the wires joined to its terminals be joined to one another, the change in the circuit is equivalent to making C infinite, and the impedance then becomes $\sqrt{R^2 + \omega^2 L^2}$.

We shall, when there is no danger of confusion or misunderstanding, speak of the complex quantities E ($\equiv E_0\ e^{i\omega t}$), V and I as the electromotive force, potential difference and current respectively ; and we shall speak of the complex quantity, $R + i(\omega L - 1/\omega C)$, or $R + i\omega L$, in the case where $C = \infty$, as the resistance.[1]

§ 23·9. The Transformer

We shall now study the case of two circuits, each having resistance and self-inductance in series, but no condenser (i.e. $C = \infty$). In one of the circuits, the primary one, we shall suppose an impressed sinusoidal E.M.F. like **(23·7)**. In the secondary circuit we shall suppose no impressed E.M.F. except such as is due to the mutual inductance between the two circuits.

---

[1] Whenever it becomes necessary to distinguish between the **real** quantity, R, usually termed ' resistance ', and the complex quantity, $R + i(\omega L - 1/\omega C)$, to which it is often convenient to apply the term ' resistance ', we shall call the former **Ohmic Resistance.**

The appropriate differential equations are easily seen to be :

$$\left. \begin{array}{l} L'\dfrac{dI'}{dt} + M\dfrac{dI''}{dt} + R'I' = \mathrm{E}_0 e^{i\omega t}, \\[2mm] M\dfrac{dI'}{dt} + L''\dfrac{dI''}{dt} + R''I'' = 0, \end{array} \right\} \qquad . \quad . \quad \textbf{(23·9)}$$

in which L', I', R' refer to the primary circuit, while L'', I'', R'' refer to the secondary circuit. M is the coefficient of mutual inductance of the circuits, and with suitable choice of the positive direction of the current in the secondary, it will be a positive constant.

A particular solution of **(23·9)** will be :

$$\left. \begin{array}{l} I' = A e^{i\omega t}, \\[1mm] I'' = B e^{i\omega t}, \end{array} \right\} \qquad . \quad . \quad . \quad . \quad \textbf{(23·91)}$$

if suitable values be assigned to the constants, A and B. These values can be determined by substitution in **(23·9)** as follows :

$$i\omega L'A + i\omega M B + R'A = \mathrm{E}_0,$$
$$i\omega M A + i\omega L''B + R''B = 0,$$

or

$$\left. \begin{array}{l} (R' + i\omega L')A + i\omega M B = \mathrm{E}_0, \\[1mm] i\omega M A + (R'' + i\omega L'')B = 0 ; \end{array} \right\} \qquad . \quad . \quad \textbf{(23·92)}$$

whence

$$A = \mathrm{E}_0 \left\{ R' + i\omega L' + \dfrac{\omega^2 M^2}{R'' + i\omega L''} \right\},$$

or

$$A = \mathrm{E}_0 \Big/ \left\{ \left(R' + \dfrac{\omega^2 M^2 R''}{R''^2 + \omega^2 L''^2} \right) + i\omega \left(L' - \dfrac{\omega^2 M^2 L''}{R''^2 + \omega^2 L''^2} \right) \right\} ; \quad . \quad \textbf{(23·93)}$$

while

$$B = - \dfrac{i\omega \, M}{R'' + i\omega L''} A,$$

or

$$B = \dfrac{\omega M}{\sqrt{R''^2 + \omega^2 L''^2}\,|} \, e^{-i(\pi/2 + \varepsilon)} A, \quad . \quad . \quad \textbf{(23·94)}$$

in which

$$\tan \, \varepsilon = \omega L''/R''.$$

If the secondary were absent, or $M = 0$, we should have for A,

$$A = \mathrm{E}_0 / \{ R' + i\omega L' \}.$$

so that the consequences of the presence of the secondary circuit,

for the primary one, are equivalent to an *increase* in *its resistance* equal to

$$\frac{\omega^2 M^2 R''}{R''^2 + \omega^2 L''^2},$$

together with a *diminution* of *its self-inductance* equal to

$$\frac{\omega^2 M^2 L''}{R''^2 + \omega^2 L''^2},$$

and a consequent *increase of its power factor*.

Equation **(23·94)** shows that the current in the secondary circuit lags in phase behind that in the primary circuit by an amount which lies between $\pi/2$ and π. In fact we may write for the current in the primary :

$$I_1' = \frac{E_0}{\sqrt{\rho^2 + \omega^2 \lambda^2}\,|} \cos (\omega t - \phi),$$

where $\tan \phi = \omega \lambda / \rho,$

and ρ and λ are respectively the apparent resistance and apparent self-inductance of the primary, i.e.

$$\rho = R' + \omega^2 M^2 R'' / (R''^2 + \omega^2 L''^2),$$
$$\lambda = L' - \omega^2 M^2 L'' / (R''^2 + \omega^2 L''^2).$$

The current in the secondary is

$$I_1'' = \frac{\omega M E_0}{\sqrt{\rho^2 + \omega^2 \lambda^2}\,|\,\sqrt{R''^2 + \omega^2 L''^2}\,|} \cos (\omega t - \phi - \pi/2 - \varepsilon).$$

The reader can easily verify that the *force of attraction* between the two circuits is equal to

$$K I_1' I_1'',$$

where K is a *positive* constant. Hence the *average force of attraction* is equal to the product of a *positive* constant and the integral :

$$\frac{\omega}{2\pi} \int_0^{2\pi/\omega} \cos (\omega t - \phi) \cos (\omega t - \phi - \pi/2 - \varepsilon) dt.$$

The value of the integral is

$$\tfrac{1}{2} \cos (\pi/2 + \varepsilon),$$

and is consequently *negative*. Repulsion therefore predominates

between the two circuits. This is the explanation of the classical experiment of Elihu Thomson.

We have tacitly assumed in the foregoing theory of the simple transformer that we are justified in adopting the particular solution described by **(23·91)** and the succeeding formulae. It is easy to show that the complementary parts of the solution, involving two arbitrary constants, rapidly approach zero, and so leave us with the particular solution we have been using. To get this complementary part of the solution we replace the right member of **(23·9)** by zero, and try to get a solution of the resulting equations, which we may term the modified equations, of the form :

$$I' = ae^{\lambda t}, \atop I'' = be^{\lambda t}, \qquad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \textbf{(23·95)}$$

a and b being constants. On substitution in the modified equations we obtain :

$$(\lambda L' + R')a + \lambda Mb = 0, \atop \lambda Ma + (\lambda L'' + R'')b = 0. \qquad \cdot \quad \cdot \quad \textbf{(23·96)}$$

If a and b are not to vanish, we must have :

$$\begin{vmatrix} \lambda L' + R', & \lambda M \\ \lambda M, & \lambda L'' + R'' \end{vmatrix} = 0. \quad \cdot \quad \cdot \quad \textbf{(23·961)}$$

Thus we learn that there are two possible values of λ, say λ_1 and λ_2. It is easy to see that, if both of these values are real, they must also be negative, and if complex they have real parts which are negative.[1] Imagine one of these roots, λ_1, to be substituted in **(23·95)** and call the amplitudes a_1 and b_1 to distinguish them from the corresponding amplitudes when the root, λ_2, is used. We learn from **(23·96)** that

$$a_1/b_1 = -\lambda_1 M/(\lambda_1 L' + R')$$
$$= -(\lambda_1 L'' + R'')/\lambda_1 M.$$

Only one of the constants a_1 and b_1 is therefore arbitrary, and the solution **(23·95)** is not the general one. If, however, we use the root, λ_2, we find another solution with one arbitrary constant. We combine these two solutions and thus obtain :

$$I' = a_1 e^{\lambda_1 t} + a_2 e^{\lambda_2 t}, \atop I'' = b_1 e^{\lambda_1 t} + b_2 e^{\lambda_2 t}, \qquad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \textbf{(23·97)}$$

---

[1] The coefficients of λ^2, λ and the remaining term in the quadratic equation **(23·961)** have all the same sign, since $L'L'' - M^2$ is positive.

in which the constants a_1 and a_2 may be arbitrary, b_1 and b_2 being determined from them by means of **(23·96)**. Equations **(23·97)** thus represent the general solution of the modified equation, and if we add them to the particular solution of **(23·9)** we obtain its general solution. The property of the λ's mentioned above ensures that the expressions **(23·97)** will asymptotically approach zero as t increases.

§ 24. EXTENSION OF KIRCHHOFF'S LAWS

Kirchhoff's laws **(§ 21·7)** constitute a comparatively simple implement for solving problems associated with networks of conductors in which constant E.M.F.s are seated. They reduce such problems to that of solving sets of linear algebraic equations. We can extend these laws in a simple way to networks the branches of which have ohmic resistance, self-inductance and capacity in series, including of course cases where, for example, the resistance or the self-inductance, or both might be vanishingly small, or where the capacity might be infinite (condenser absent). The extension only applies when the impressed E.M.F.s in the branches vary with time in a sinusoidal or simple harmonic way, and when all the impressed E.M.F.s have the same frequency.[1]
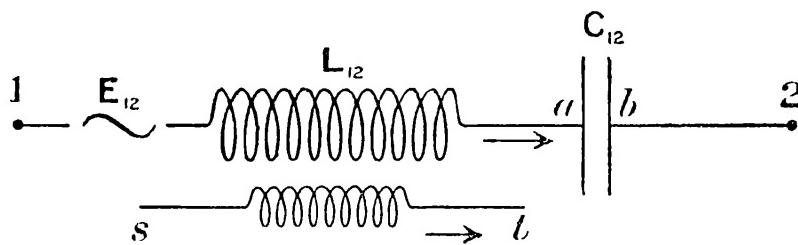
Fig. 24

These extended laws reduce the mathematical calculations associated with such network problems to the comparatively simple one of solving linear algebraic equations : in fact, any such problem becomes, mathematically, identical with a corresponding one for a network in which constant E.M.F.s are seated, and in which only ohmic resistance is involved.

Let us consider first a typical branch, 12 in Fig. 24, of such a network, the positive direction being from 1 to 2. Seated in it is an impressed E.M.F., which we may call E_{12}, directed from 1 to 2, and which may be expressed as the product of a positive constant and $\cos \omega t$. Superposed on this E.M.F. may be one or more E.M.F.s due to the inductive influence of other branches of the network, such as st in Fig. 24. Further, the branch 12 may

---

[1] We shall not exclude the possibility of mutual induction between one branch and another.

contain self-inductance, L_{12}, and capacity, C_{12}. The differential equation for the branch has therefore the form :

$$\mathbf{E}_{12} - M_{st}\frac{dI_{st}}{dt} - L_{12}\frac{dI_{12}}{dt} + V_1 - V_a + V_b - V_2 = R_{12}I_{12}, \quad (24)$$

in which $V_a - V_b$ will be replaced, at a subsequent stage, by Q_{12}/C_{12}. We must make use of the device explained in **§ 23·7**, and give $\mathbf{E}_{12}$ the form :

$$\mathbf{E}_{12} = \mathbf{A}_{12}e^{i\omega t},$$

where $\mathbf{A}_{12}$ is a real and positive constant. On differentiating (24) we obtain the equation

$$iw\mathbf{A}_{12}e^{i\omega t} - M_{st}\frac{d^2 I_s}{dt^2} - L_{12}\frac{d^2 I_{12}}{dt^2} - \frac{1}{C_{12}}I_{12}$$
$$+ \frac{dV_1}{dt} - \frac{dV_2}{dt} = R_{12}\frac{dI_{12}}{dt}. \quad \cdot \quad \cdot \quad \cdot \quad (24\cdot01)$$

When we add together all the equations for a mesh, or closed loop, in the network, the terms dV/dt annul one another, as in **§ 21·7**, and we obtain an equation which has the form :

$$e^{i\omega t}\Sigma i\omega A_{\alpha\beta} = \Sigma M_{st}\frac{d^2 I_{st}}{dt^2} + \Sigma L_{\alpha\beta}\frac{d^2 I_{\alpha\beta}}{dt^2} + \Sigma R_{\alpha\beta}\frac{dI_{\alpha\beta}}{dt} + \Sigma c_{\alpha\beta}I_{\alpha\beta}. \quad (24\cdot02)$$

In this equation the $\alpha\beta$ refer to the branches of the closed loop, the st refer to those parts of the network which produce induced E.M.F.s in it, and $c_{\alpha\beta}$ is an abbreviation for $1/C_{\alpha\beta}$. We have a differential equation like (24·02) for every closed loop in the network, and it is easy, by a simple extension of the method indicated in **§ 23·9**, to obtain the general solution of the whole set of equations. For the reasons explained in **§§ 23·7** and **23·9** the complementary part of this general solution soon becomes evanescent, and we are left with expressions for the currents of the form :

$$I_{12} = \mathbf{I}_{12}e^{i\omega t}, \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad (24\cdot03)$$

where $\mathbf{I}_{12}$ is a constant, and not necessarily real. We can infer at once that the potential, V, at any point on the network has a similar form, namely

$$V = \mathbf{V}e^{i\omega t}. \quad \cdot \quad \cdot \quad \cdot \quad (24\cdot031)$$

Similarly, the charge Q on any condenser is

$$Q_{12} = \mathbf{Q}_{12}e^{i\omega t},$$

and consequently

$$I_{12} = i\omega Q_{12}. \quad \cdot \quad \cdot \quad \cdot \quad (24\cdot032)$$

It is now clear that we may write instead of (24)

$$\mathbf{E}_{12} - i\omega M_{st}I_{st} + V_1 - V_2 = \left\{ \mathbf{R}_{12} + i\left(\omega L_{12} - \frac{1}{\omega C_{12}}\right) \right\} I_{12}, \quad \textbf{(24·04)}$$

remembering that $Q_{12}/C_{12} = V_a - V_b$ and making use of (24·032), or finally

$$\mathbf{E}_{12} + V_1 - V_2 = \mathbf{R}_{12}I_{12}, \quad \quad \textbf{(24·05)}$$

where $\mathbf{E}_{12} \equiv \mathbf{E}_{12} - i\omega M_{st}I_{st},$

and $\mathbf{R}_{12} = \mathbf{R}_{12} + i\left(\omega L_{12} - \frac{1}{\omega C_{12}}\right).$

The variable part of every term in (24·05) is simply contained in the common factor $e^{i\omega t}$, so that when we divide by it we are left with an equation in every way comparable with (21·71), except for the fact that the quantities involved are in general complex. It follows that for any mesh in a network, in every part of which the impressed E.M.F. is sinusoidal and of the same angular frequency, ω, everywhere, we may adopt equations analogous to (21·7), namely :

$$\varSigma\mathbf{E} = \varSigma\mathbf{R}I, \quad \quad \quad \quad \textbf{(24·06)}$$

in which $\mathbf{E}$, $\mathbf{R}$ and I are the complex quantities, the meaning of which has already been explained.

We have been assuming, while dealing with alternating currents, that the conduction current along a wire has the same strength at all cross-sections of the wire. This is not of course strictly true, since the variation of the electric field will give rise to displacement currents in the surrounding air or insulating medium. These displacement currents are, however, inappreciable except through the condenser itself (or through the condensers). This becomes obvious when we remember that for a given rate of change of the field intensity the associated displacement current is proportional to the area through which it is passing ; and the area of the surface of the conducting wire, or wires, in the circuits we are studying is very minute compared with that of the condenser plates. We may therefore adopt Kirchhoff's first law here, though strictly speaking it is only approximate. In the widest sense, namely when applied to currents of every description—conduction, convection or displacement currents—Kirchhoff's first law is of course always true, since it expresses the fundamental law of **conservation of electricity.**

§ 24·1. ALTERNATING CURRENTS IN NETWORKS

By way of illustration we shall apply the extended Kirchhoff laws to some simple cases ; firstly, to Anderson's method of determining a self-inductance in terms of a capacity and resistances. This is illustrated in Fig. 24·1. Fundamentally it is a Wheatstone bridge, the four arms of which have the resistances, in the ordinary ohmic sense, P, Q, R and S. The arms P, Q and R have negligible inductance. The portion, r, is also non-inductive and has an ohmic resistance, r. The branch, C, has a condenser of capacity, C, in series with a negligible inductance and negligible ohmic resistance. The branch, S, has ohmic resistance, S, and a self-inductance, L, in series. E represents a sinusoidal E.M.F. of angular frequency, ω. Finally, G is a telephone (or galvanometer when this is needed). The method of experiment is to

FIG. 24·1

adjust so that the current through G shall be zero. The arrows indicate the directions selected as positive. Applying the extended second law of Kirchhoff to the three meshes C, r, P ; R, G, C and S, Q, r, G respectively, and remembering that the current in G and the E.M.F. round each mesh is zero, we get :

$$-\frac{i}{\omega C}I_C + rI_r - PI_P = 0,$$

$$RI_R + \frac{i}{\omega C}I_C \qquad = 0,$$

$$(S + i\omega L)I_S - QI_Q - rI_r = 0.$$

The first law gives us in addition :

$$I_P + I_r = I_Q,$$

$$I_C = I_r,$$

$$I_R = I_S.$$

When we eliminate the currents from these equations, we are left with :

$$S + i\omega L - i\omega CRQ - i\omega CRr - i\omega Cr\frac{QR}{P} - \frac{QR}{P} = 0. \quad \textbf{(24·1)}$$

From this equation we get by equating the real parts :
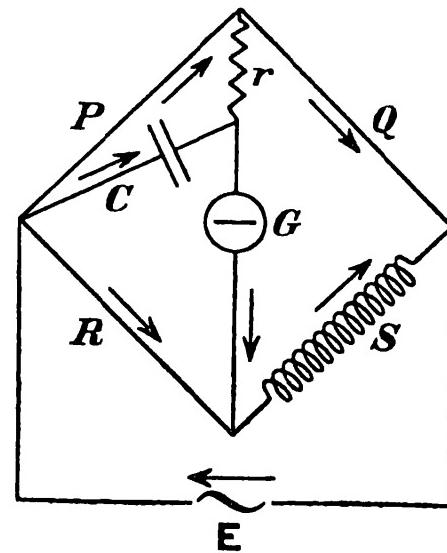
$$S - QR/P = 0, \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \textbf{(24·11)}$$

which means that the bridge must be *balanced for steady currents*. From the purely imaginary part of **(24·1)**, remembering **(24·11)**, we get

$$L/C = QR + r(R + S). \quad . \quad . \quad . \quad \textbf{(24·12)}$$

It will be observed that the frequency, ω, cancels out, and that, consequently, the formula **(24·12)** applies for any frequency. Furthermore, the linear character of the differential equations, on which our theory is based, leads to the consequence that **(24·12)** will hold for E.M.F.s and currents which are superpositions of sinusoidal E.M.F.s and currents, and Fourier's expansion, or Fourier's theorem, enables us to represent any varying quantity as such a superposition. It therefore follows that we may apply any sort of periodic E.M.F. in Anderson's (or Maxwell's, *infra*) bridge ; or even one that is not periodic at all.

The special case where r is made zero represents the original form of this experiment, and is due, like several other similar experimental methods, to Clerk Maxwell. The original method has the disadvantage that both PS and L/C must be equal to the same quantity, QR, to produce silence in the telephone. Anderson's improvement consists in giving us liberty to leave the adjustment, $PS = QR$, untouched after it has once been made, since the requirement **(24·12)** can be secured (if QR or PS are initially made small enough) by changing r only.

Fig. 24·11

As a further illustration let us consider the method, illustrated in Fig. 24·11, of determining a self-inductance in terms of a mutual inductance. The former, L, is the self-inductance of the coil b, while the latter, M, is the mutual inductance of the coils a and b. Applying the extended Kirchhoff second law to the meshes R, G, P ; S, Q, G and P, Q, r respectively, we obtain :

$$i\omega MI = (R + i\omega L)I_R - PI_P,$$

since $i\omega MI$ functions here as the impressed E.M.F. in the arm R ;

$$0 = SI_S - QI_Q,$$
$$0 = PI_P + QI_Q - rI_r.$$
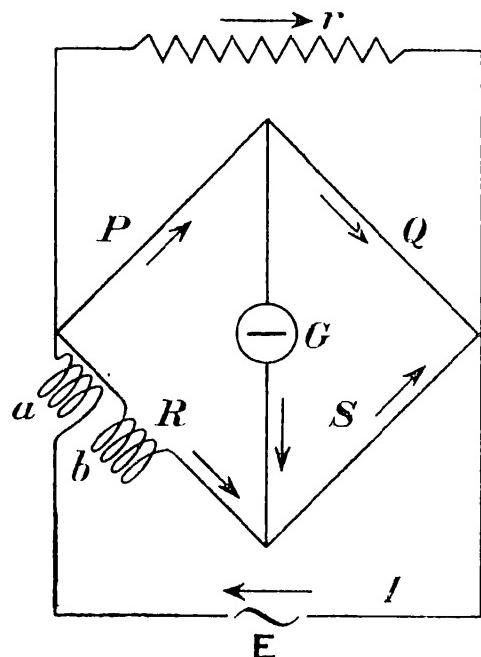
The first law gives us :

$$I = I_R + I_P + I_r.$$
$$I_P = I_Q,$$
$$I_R = I_S.$$

When we eliminate the currents we are left with :

$$i\omega M\left\{1 + \frac{S}{Q} + \frac{1}{r}\left(\frac{PS}{Q} + S\right)\right\} = R + i\omega L - \frac{PS}{Q}.$$

Equating the real parts of this yields

$$R = PS/Q, \quad . \quad . \quad . \quad . \quad . \quad \textbf{(24·13)}$$

and we get by equating the imaginary parts :

$$M\left\{1 + \frac{S}{Q} + \frac{R + S}{r}\right\} = L. \quad . \quad . \quad . \quad \textbf{(24·14)}$$

The last formula indicates that L must necessarily exceed M. Just as in the case of Anderson's bridge, and for the same reasons, we may here use *any* periodic E.M.F., not merely a sinusoidal one, a telephone being placed in the arm G. Or we may use any cell and key in conjunction with a ballistic galvanometer. Furthermore, this method, like Anderson's, is an improvement of an older method due to Clerk Maxwell. In Maxwell's original method, the arm, r, is absent ; i.e. r is infinite, and the formula **(24·14)** becomes

$$M\left\{1 + \frac{S}{Q}\right\} = L. \quad . \quad . \quad . \quad \textbf{(24·15)}$$

The later method has the advantage that the condition **(24·14)** can be satisfied (by suitably adjusting r) without necessitating any change in the original adjustment for **(24·13)**.

§ 24·2. A DIGRESSION

There are certain bridge methods resembling those described in the foregoing section, which however are founded on a somewhat different theoretical basis. In these latter a constant E.M.F. is suddenly impressed in the system (or removed from it) by the use of a cell and key ; so that the initial state of affairs is one of zero current everywhere, followed by a rapid rise to final steady values (or initial steady values followed by a rapid drop to zero). Such a device can be used in the methods already described, if a ballistic galvanometer be placed in the arm G.

The device will succeed however in cases like those about to be described, in which the current in the arm, G, cannot be made equal to zero until the steady state has been reached, provided the *algebraic quantity* of electricity discharged (in the sense of the arrow, for example) through the galvanometer is zero. *It is necessary* that the moving system of the galvanometer should have a big moment of inertia in order that no appreciable motion may be set up during the short time or times when the current may actually differ from zero. The theory of the preceding section fails here, since it is, of course, directed to secure *zero current* (*and not merely zero algebraic quantity*).

We shall deal first with Rimington's method, which closely resembles Anderson's. The reader will easily verify by the method used in the last section that the current in the branch, G, cannot in general be made equal to zero.

Fig. 24·2

In order to deal with *quantity*, as distinct from current strength, we apply Kirchhoff's laws in their original form. Rimington's bridge is shown diagrammatically in Fig. 24·2. As usual there are four arms with resistances P, Q, R, S. The arm S is divided into two portions, the resistances of which are s and s', so that $S = s + s'$. A condenser of capacity C is connected on one side to the junction of s and s', and on the other to the point where the current emerges from (or enters) the bridge system.

A self-inductance, L, is contained in the arm, P. On closing the key, K, the current rises rapidly from zero to steady final values in the various branches of the network. Taking the mesh, P, G, R, there is an E.M.F. through P, in the direction of the arrow, equal to $- L\dfrac{dI_P}{dt}$ and consequently

$$- L\frac{dI_P}{dt} = PI_P - GI_G - RI_R.$$

On multiplying by dt and integrating over the period of time during which the currents rise from zero to what is practically the final value, we get

$$- LI_P = P\mathfrak{Q}_P - R\mathfrak{Q}_R, \quad . \quad . \quad . \quad . \quad \textbf{(24·2)}$$

where I_P now represents the final steady value of the current in
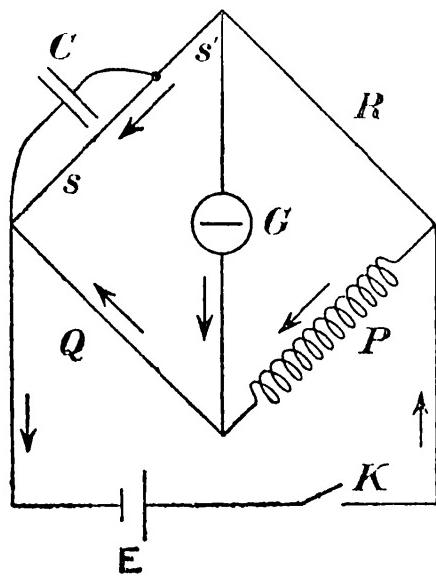
the branch, P, and $\mathfrak{Q}_P$ and $\mathfrak{Q}_R$ are the respective quantities of electricity discharged, in the senses of the arrows, through P and R respectively. By hypothesis $\mathfrak{Q}_G$ the algebraic quantity discharged through the galvanometer is zero.

Turning to the mesh Q, G, s', s, we find, since the E.M.F. in it is zero,

$$0 = -QI_Q - GI_G + s'I_{s'} + sI_s.$$

Integrating, as before, we get :

$$0 = -Q\mathfrak{Q}_Q + s'\mathfrak{Q}_{s'} + s\mathfrak{Q}_s. \quad . \quad . \quad . \quad (24·21)$$

Now $\mathfrak{Q}_{s'}$ must be equal to $\mathfrak{Q}_R$ since $\mathfrak{Q}_G$ is zero, and for the same reason $\mathfrak{Q}_Q = \mathfrak{Q}_P$. Further, when the final steady state has been reached, the p.d. between the condenser terminals is sI_R, and the quantity which has entered the condenser is sCI_R. Hence

$$\mathfrak{Q}_s = \mathfrak{Q}_R - sCI_R.$$

Substituting in (24·21) we get :

$$0 = -Q\mathfrak{Q}_P + s'\mathfrak{Q}_R + s(\mathfrak{Q}_R - sCI_R)$$

or

$$0 = S\mathfrak{Q}_R - Q\mathfrak{Q}_P - s^2CI_R. \quad . \quad . \quad . \quad . \quad (24·22)$$

If now we suppose the bridge to be adjusted so that no current flows through G when the final steady state is reached ; i.e. $PS = QR$, then (24·22) becomes :

$$s^2CI_R = \frac{QR}{P}\mathfrak{Q}_R - Q\mathfrak{Q}_P,$$

or

$$\frac{s^2CPI_R}{Q} = R\mathfrak{Q}_R - P\mathfrak{Q}_P. \quad . \quad . \quad (24·23)$$

On comparing this with (24·2) we see that

$$LI_P = \frac{s^2CPI_R}{Q},$$

and obviously

$$I_P/I_R = R/P,$$

consequently

$$\frac{L}{C} = \frac{s^2P}{S}. \quad . \quad . \quad . \quad . \quad . \quad (24·24)$$

When the condenser terminals are joined to the ends of the arm S, i.e. when $s = S$, (24·24) becomes :

$$L/C = SP = QR,$$
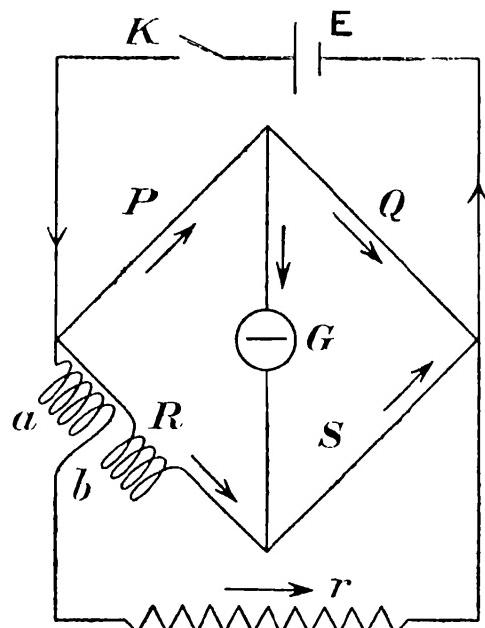
Fig. 24·21

and we see that Maxwell's bridge is a special case of Rimington's bridge, as we have seen it to be a special case of Anderson's.

Another example is the method of determining a mutual inductance in terms of a self-inductance illustrated by Fig. 24·21. The self-inductance, L, is in the arm, R, and M is the mutual inductance of the coils a and b. Applying Kirchhoff's second law to the meshes R, G, P and S, Q, G respectively, we get :

$$- L\frac{dI_R}{dt} + M\frac{dI_r}{dt} = RI_R - GI_G - PI_P,$$

$$0 = SI_S - QI_Q + GI_G.$$

Integrating as before we find, since $\mathfrak{Q}_G = 0$,

$$- LI_R + MI_r = R\mathfrak{Q}_R - P\mathfrak{Q}_P,$$

$$0 = S\mathfrak{Q}_S - Q\mathfrak{Q}_Q.$$

If now we arrange that $PS = QR$, and remember that $\mathfrak{Q}_S = \mathfrak{Q}_R$ and $\mathfrak{Q}_Q = \mathfrak{Q}_P$, the second equation becomes :

$$0 = R\mathfrak{Q}_R - P\mathfrak{Q}_P.$$

Hence the first equation becomes :

$$LI_R = MI_r. \quad . \quad . \quad . \quad . \quad \textbf{(24·25)}$$

On applying Kirchhoff's second law to the steady state in the mesh R, S, r, we get :

$$0 = RI_R + SI_S - rI_r$$

and, since the bridge is balanced for steady currents,

$$I_R = I_S,$$

hence

$$I_r = \frac{R + S}{r}I_R.$$

Substituting this in **(24·25)** we get finally :

$$L = M\left(\frac{R + S}{r}\right). \quad . \quad . \quad . \quad \textbf{(24·26)}$$

In this case L may be either greater or less than M as the formula **(24·26)** indicates.

Carey Foster's method of determining a mutual inductance may be given as a final illustration of the ballistic or integrated type of balance.   It will be understood by reference to Fig. 24·22, in which the two coils of mutual inductance, M, are indicated by 1 and 2. A cell, E, a ballistic galvanometer, G, and a condenser of capacity, C, are arranged as shown.   When the key, K, is closed a current starts in coil 1 and in consequence an E.M.F. is induced in 2.   The throw of the galvanometer can be reduced to zero by suitably adjusting the resistances, r_1 and r_2, in two non-inductive branches shown in the figure. By applying Kirchhoff's second law to the mesh containing r_2, G and the coil 2, we get :

Fig. 24·22

$$M\frac{dI_1}{dt} - L_2\frac{dI_G}{dt} = GI_G + r_2I_2,$$

in which the meanings of the various symbols are obvious.   On integrating we obtain :

$$MI_1 = r_2\mathcal{Q}, \qquad . \quad . \quad . \quad . \quad . \quad \textbf{(24·27)}$$

where $\mathcal{Q}$ means the quantity of electricity discharged along r_2. It is in fact the quantity of electricity that has entered the condenser, as $\mathcal{Q}_G$ is zero.   When the final state is reached, the p.d. between the terminals of the condenser is r_1I_1, and hence its charge must be r_1I_1C ;   consequently

$$\mathcal{Q} = r_1I_1C,$$

and on substituting in **(24·27)** and dividing through by I_1, we arrive at the formula :

$$M = r_1r_2C. \qquad . \quad . \quad . \quad . \quad . \quad \textbf{(24·28)}$$

§ 24·29.   Propagation of an Electrical Disturbance along a Cable

We shall now study what happens in a long wire or cable after the removal of some temporary E.M.F. which has produced
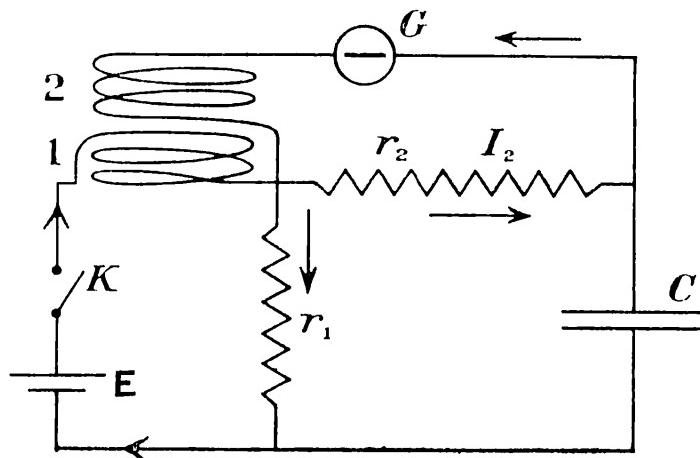
a local variation of potential in it. We shall suppose that the self-inductance, capacity and resistance per unit length of the cable are constants, i.e. have the same value at all parts of the cable, and represent them by L, C and R respectively. The current passing the cross-section at x (Fig. 24·29) and the potential at that point may be represented by I and V respec-

tively, the potential of the surrounding neighbourhood being taken as zero. We shall not assume the cable to be
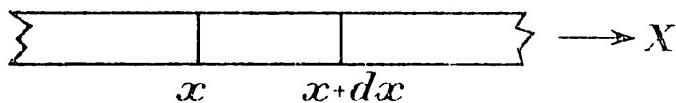
x $x+dx$

Fig. 24·29

perfectly insulated, so that in general a current will flow between it and the surrounding neighbourhood. Let us suppose this current to be SV per unit length of the cable, where S is a constant : the **leakance** per unit length. We may write down two expressions for the algebraic quantity of electricity entering the element dx (Fig. 24·29) during the interval dt, namely :

$$Cdx dV,$$

obtained by multiplying the corresponding mean increase of potential by the capacity of the element ; and

$$Idt - (I + dI)dt - SdxVdt,$$

obtained by considering the currents flowing into and out of the element. Therefore

$$Cdx dV = - dIdt - SVdxdt.$$

In this equation dV means the rise in potential at a definite place (x constant) during the time dt. It is therefore equal to $\frac{\partial V}{\partial t}dt$. Similarly, $dI = \frac{\partial I}{\partial x}dx$, since it means the amount by which the current at $x + dx$ exceeds that at x *at the same time*. Making these substitutions in the last equation we get :

$$C\frac{\partial V}{\partial t} + SV = - \frac{\partial I}{\partial x}. \quad . \quad . \quad . \quad (24\cdot29)$$

We obtain a second equation by considering the electromotive force along the element dx and the potential difference between its ends, namely :

$$- Ldx\frac{\partial I}{\partial t} - dV = RdxI,$$

where dV now means $dx\partial V/\partial x$.

Therefore

$$L\frac{\partial I}{\partial t} + RI = - \frac{\partial V}{\partial x}. \quad . \quad . \quad (24\cdot291)$$

From these two equations we may eliminate V and its differential quotients, or, alternatively, I and its differential quotients. Adopting the latter alternative, we differentiate **(24·29)** partially with respect to t, and **(24·291)** partially with respect to x, and then eliminate $\partial^2 I/\partial x \partial t$ from the resulting equations, thus obtaining :

$$LC\frac{\partial^2 V}{\partial t^2} + LS\frac{\partial V}{\partial t} = R\frac{\partial I}{\partial x} + \frac{\partial^2 V}{\partial x^2}.$$

Finally, we get rid of $\partial I/\partial x$ in this last equation by the aid of **(24·29)** and so arrive at the following differential equation for V :

$$LC\frac{\partial^2 V}{\partial t^2} + (LS + RC)\frac{\partial V}{\partial t} + RSV = \frac{\partial^2 V}{\partial x^2}. \quad . \quad \textbf{(24·292)}$$

When LS happens to be equal to RC the equation is identical in form with **(9·12)**, as we see when we denote LC by $1/u^2$ and RS by a^2 ; and a solution of it is :

$$V = e^{-\sqrt{RS}|x} f\left(t/\sqrt{LC} \mid - x\right), \quad . \quad \textbf{(24·293)}$$

where f is an arbitrary function. Subject therefore to the condition that $LS = RC$, an electrical disturbance of any form, since f is arbitrary (see **§ 9·1**), will be propagated without change in form (without distortion we may say) and with a velocity, namely $1/\sqrt{LC}\,|$, which depends only on the constants of the cable. The amplitude of the disturbance will diminish exponentially, the rate of attenuation depending on the product RS.

When $LS \neq RC$ we can easily obtain particular solutions of **(24·292)** by Bernoulli's method. If we substitute for V the product

$$e^{i\omega t} X,$$

where X is some function of x only, we can determine the form of X, and we thus find a solution which represents a damped harmonic wave. In this case, however, the velocity of propagation is found to *depend on the frequency*, ω. Since an arbitrary disturbance can be represented by Fourier's methods as a superposition of harmonic components, and since each of these has its characteristic velocity of propagation, we realize that in general the disturbance in the cable cannot be propagated without change in form, or distortion. The subject of this section has an important practical bearing : the condition $LS = RC$ being essential for successful telephony, especially when the cable is a submarine one.[1]

---

[1] O. Heaviside : *Electromagnetic Theory*, Vol. I, p. 368.

§ 24·3. Practical Units

Practical considerations have led to the introduction and common use of certain units, called **practical units,** to which reference has been made in **§ 22·1.** The practical unit of work and energy is the **joule,** which is defined to be 10^7 **ergs.** The practical unit of current strength is the **ampere,** defined as 1/10 of an ordinary electromagnetic unit of current strength, and consequently the **coulomb** (the practical unit of quantity) may be defined as the quantity transported by an **ampere** during one second. The practical unit of E.M.F. and P.D. is the **volt,** which is defined by the work, expressed in **joules,** done on the unit quantity expressed in **coulombs.** The practical unit of **capacity** is of course one **coulomb** per **volt,** and is termed the **farad.** This unit is inconveniently large for laboratory purposes, and standards of capacity are usually expressed in **microfarads.** One **microfarad** is equal to 10^{-6} **farad.** The practical unit of **resistance** is the **ohm,** defined of course as **one volt** per **ampere.** The practical unit of **inductance** is the **henry** and is defined (whether we are dealing with self or mutual inductance) to be the inductance of a coil or circuit in which an induced E.M.F of **one volt** is occasioned by a rate of change of current equal to **one ampere per second.**

The numerical relationships contained in the definitions just given enable us to evaluate those which hold between any of the various electrical and other units of whatever type. First of all we start from

$$V_v = W_j/Q_c,$$

where V_v means a P.D. expressed in volts, W_j means work expressed in joules and Q_c means quantity of electricity expressed in coulombs. Therefore

$$V_v = 10^{-7} W_e/10 Q_{e.m.},$$

where W_e means the same amount of work as does W_j, but expressed in ergs, and $Q_{e.m.}$ means the same quantity of electricity as before, but expressed in ordinary E.M. units. Hence

$$10^8 V_v = W_e/Q_{e.m.},$$

or

$$\boxed{10^8 V_v = V_{e.m.}}$$

Hence

$$\boxed{\textbf{one volt} = 10^8 \times \textbf{E.M. unit of P.D.}}$$

Consider next two equal point charges separated by the unit distance. We have

$$\frac{Q^2{}_{e.s.}}{K_{e.s.}} = \frac{Q^2{}_{e.m.}}{K_{e.m.}},$$

since either of these expressions represents the force (in dynes) between the two charges. Therefore

$$\frac{Q_{e.m.}}{Q_{e.s.}} = \sqrt{\frac{K_{e.m.}}{K_{e.s.}}}.$$

These ratios are **not the ratios of the units themselves, but of the measures of the respective magnitudes, quantity and dielectric constant, in terms of the ordinary E.M. and E.S. units respectively.** We are thus led to

$$\frac{Q_{e.m.}}{Q_{e.s.}} = \frac{1}{c},$$

(§ 22·1), or

$$\frac{Q_{e.s.}}{Q_{e.m.}} = 3 \times 10^{10}.$$

It follows that :

$$\left\{\begin{array}{l}\textbf{Ordinary E.M. unit}\\ \textbf{of quantity}\end{array}\right\} = \textbf{3} \times \textbf{10}^{\textbf{10}} \times \left\{\begin{array}{l}\textbf{Ordinary E.S. unit}\\ \textbf{of quantity}\end{array}\right\}$$

and further that :

$$\left\{\begin{array}{l}\textbf{Ordinary E.S. unit of}\\ \textbf{potential difference}\end{array}\right\} = \textbf{300 volts.}$$

The definition of the coulomb implies that the **second** is the practical unit of time ; but the system of practical units has never been completed. One way in which it might be extended further would be to adopt the kilogram as the practical unit of mass and the metre as the practical unit of length. The unit of force would thus be 10^5 dynes, and the choice would be consistent with the practical unit of work being 10^7 ergs. If we define dielectric constant by using for the force between two equal point charges the expression

$$F = Q^2/Kr^2,$$

the practical unit for K, with the choice given above for units of mass and length, would be 10^{-11} times the ordinary E.M.U. of dielectric constant, or 9×10^9 times the ordinary E.S.U.

§ 24·4. DIMENSIONS OF ELECTRICAL QUANTITIES

One expression for the dimensions of an electric charge has already been given **(19·16)**. If Q, K, M, L and T represent charge, dielectric constant, mass, length and time respectively,

$$Q = K^{1/2}M^{1/2}L^{3/2}T^{-1}. \qquad \textbf{(24·4)}$$

Similarly,

$$m = \mu^{1/2}M^{1/2}L^{3/2}T^{-1}, \qquad \textbf{(24·41)}$$

where m and μ mean respectively pole strength and magnetic permeability.

The dimensions of electric field intensity, E, and magnetic field intensity, H, are easily obtained by dividing the dimensions of force **(§ 19·1)** by those of charge and pole strength respectively. We thus get :

$$\left.\begin{array}{l} E = M^{1/2}L^{-1/2}T^{-1}K^{-1/2}, \\ H = M^{1/2}L^{-1/2}T^{-1}\mu^{-1/2}. \end{array}\right\} \qquad \textbf{(24·42)}$$

And difference of electrical potential and E.M.F. being measured by work per unit charge must have the dimensions :

$$V = M^{1/2}L^{1/2}T^{-1}K^{-1/2}. \qquad \textbf{(24·43)}$$

Equation **(22·3)** leads at once to the dimensional equation :

$$QT^{-1} = HL,$$

or **(24·42)**

$$QT^{-1} = M^{1/2}L^{1/2}T^{-1}\mu^{-1/2},$$

whence

$$Q = M^{1/2}L^{1/2}\mu^{-1/2}. \qquad \textbf{(24·44)}$$

If we equate the expressions **(24·4)** and **(24·44)** for the dimensions of Q, we get :

$$\mu^{-1/2}K^{-1/2} = LT^{-1}, \qquad \textbf{(24·45)}$$

a result which is verified by the fact that the velocity of electromagnetic waves is equal to $a/\sqrt{\mu K}\,|$, a being a numerical constant **(§§ 19·1 and 24·8)**. The reader will easily be able to find expressions for the dimensions of other electrical or magnetic quantities by writing down the dimensional equations which emerge from the formulae connecting the measures of these quantities with those of the fundamental quantities. We

shall mention only those of inductance, λ, capacity, κ, and resistance, R.

$$\left.\begin{aligned} \lambda &= \mu\text{L}, \\ \kappa &= \text{KL}, \\ \text{R} &= \mu\text{LT}^{-1}. \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad \textbf{(24·46)}$$

It should be emphasized that in the foregoing account of the dimensions of electrical quantities no mention has been made of any particular choice of units, electrostatic, electromagnetic or other. The expressions for the dimensions have nothing to do with the choice of units.

THERMODYNAMICAL ASPECTS OF ELECTRIC CURRENTS

§ 24·5. Electromotive Force as a Function of the Temperature

THE E.M.F. of a cell is a function of its temperature. It is convenient and instructive to study the more general possibility of a circuit in which the impressed E.M.F. is a function of the two variables : temperature, T, and the algebraic quantity of electricity, q, which has passed round the circuit. The circuit constitutes a thermodynamical system to which equation (15·6) must apply ; so that

$$dQ = dU + dW,$$

where dQ represents energy supplied from outside in the form of heat, dU is the associated increment of the internal energy of the system and dW the work which might be done externally by a motor in the circuit. We shall suppose the moving parts of the motor to be free from friction and that it does work at an infinitely slow rate.[1] The current differs infinitesimally from zero, and if we suppose the state of the cell (or whatever it may be which provides the impressed E.M.F.), chemical or otherwise, to be uniquely determined by the temperature, T, and the quantity of electricity, q, which has passed through it ; that is, if we

---

[1] A motor may do work at zero rate in either of *two* cases : (1) the load may be just so great that the motor does not move at all ; (2) it may rotate so rapidly that the induced back E.M.F. balances the impressed E.M.F. and the current is reduced to infinitesimal dimensions. In the former case the efficiency is zero ; in the latter case (which is that we are supposing) the efficiency is unity. The relevant equation is

$$EC = RC^2 + w,$$

where E is the *impressed* E.M.F., C is the current, R the resistance and w the rate of working. The left hand represents energy supplied per unit time, while RC^2 represents heat generated per unit time. It is obvious that w vanishes when $E = RC$ (case (1)) and when $C = 0$ (case (2)). When w is a maximum (i.e. when $dw/dC = 0$) we have $E \quad 2RC$ and the efficiency, w/EC, in this case is 1/2.

assume an equation of state exists which expresses the internal
energy, the E.M.F. and other thermodynamical variables in
terms of T and q, just as in earlier studies we have found
internal energy, pressure, etc., to be functions of T and the
volume, V ; then we have a *reversible* thermodynamical system.
The work done, dW, may now be written $\mathrm{E}dq$, the E.M.F. taking
the place of the pressure in (15·91) and dq that of dV.   Hence

$$dQ = dU + \mathrm{E}dq. \quad . \quad . \quad . \quad (24\text{·}5)$$

This may be written (cf. equation 16·92) :

$$dQ = \left(\frac{\partial U}{\partial T}\right)_q dT + \left\{\left(\frac{\partial U}{\partial q}\right)_T + \mathrm{E}\right\}dq,$$

and the method of § 16·9 gives at once :

$$\left(\frac{\partial U}{\partial q}\right)_T + \mathrm{E} = T\left(\frac{\partial \mathrm{E}}{\partial T}\right)_q,$$

which may be put in the form :

$$\mathrm{E} = -\left(\frac{\partial U}{\partial q}\right)_T + T\frac{d\mathrm{E}}{dT}, \quad . \quad . \quad . \quad (24\text{·}51)$$

when E depends on the temperature only.   The last equation is
due to v. Helmholtz, and is easily seen to be a special case of
the equation (17·151), which is usually called the Gibbs-
Helmholtz equation, since E represents the diminution of the
free energy of the cell reckoned per unit quantity passing through
it.   The quantity $- (\partial U/\partial q)_T$ represents the diminution, per
unit quantity of electricity passing, of the chemical energy within
the cell in consequence of the chemical changes occurring within it.

§ 24·6.   Thermoelectric Phenomena

Such phenomena as the Joulean development of heat (§ 21·9)
in the conducting parts of a circuit might quite properly be
called thermoelectric ; but the term is usually confined to certain
phenomena which appear in circuits made of two or more different
metals.   Experiment shows that, when the junctions in a circuit
consisting simply of two or more wires A, B, C . . . have not all
the same temperature, a current is generated in it, and associated
with such a current is a corresponding E.M.F., which may be
called a **thermo-electromotive force.**   This electromotive force
may be regarded as the algebraic sum of a number of parts,
some of them seated in the junctions, and the others distributed
along each metal.   We shall use the symbol $\varPi$ for the former

and K for the latter ; so that in a circuit made up of three metals, A, B and C, for example,

$$E = \Pi(AB) + \Pi(BC) + \Pi(CA) + K(A) + K(B) + K(C). \quad . \quad (24 \cdot 6)$$

The E.M.F., $\Pi(AB)$, we shall term a **Peltier E.M.F.** It is the E.M.F. seated in the junction AB and directed from A to B, and its measure is the quantity of energy (in this case energy in the form of heat) entering the junction from outside per unit quantity of electricity crossing the junction AB, from A to B. Similarly, the E.M.F. $K(A)$ means the energy supplied (in the form of heat) to the wire A per unit quantity of electricity passing through it in the direction indicated by the order of the letters A, B, C, and we shall call it a **Kelvin E.M.F.** It is a fact which can be directly demonstrated experimentally that a definite quantity of heat (positive or negative) is absorbed at a junction AB when the unit quantity of electricity crosses from A to B, and that an equal quantity is emitted when the unit quantity of electricity crosses the junction in the opposite direction. This is the quantity, Π, which is found to depend only on the pair of metals AB, and on the temperature of the junction. It is sometimes called the **Peltier coefficient** of the junction, after Peltier who discovered it. Experiment demonstrates further that the net E.M.F. in a circuit consisting of a number of metals A, B, C . . . is zero when all the *junctions* have the same temperature, however the temperature may vary along the individual metals. This would be the case if we were to assume that the Kelvin E.M.F., $K(A)$, in a metal A is expressible in the form :

$$K(T) - K(T_0), \quad . \quad . \quad . \quad . \quad . \quad (24 \cdot 61)$$

T and T_0 being the temperatures at its ends, and we shall adopt the convention that $(24 \cdot 61)$ represents the E.M.F. in the sense from the point at the temperature, T_0, to the point at the temperature, T. Consequently the Kelvin E.M.F., in the direction of rising temperature, between two points differing in temperature by dT is :

$$dK(T) = \frac{dK(T)}{dT}dT,$$

or $\qquad\qquad dK(T) = \sigma(T)dT, \quad . \quad . \quad . \quad . \quad (24 \cdot 62)$

if we represent dK/dT by σ. We can now prove that

$$E(A_1A_2) + E(A_2A_3) + E(A_3A_4) + \ . \ . \ . \ E(A_nA_1) = 0,. \quad (24 \cdot 63)$$

where $E(A_3A_4)$, for example, means the thermoelectromotive force in the circuit consisting of the two metals A_3 and A_4, the

junctions having the temperatures T and T_0, and the E.M.F. having the sense from A_3 to A_4 across the junction at the temperature, T. This must be the case since the algebraic sum of the Πs corresponding to the temperature, T, is zero and that corresponding to the temperature, T_0, is likewise zero, while the Kelvin E.M.F.s vanish in each of these cases in virtue of **(24·61)**. Another consequence of the facts and hypotheses mentioned above is that the E.M.F. in a circuit consisting of n metals, A_1, A_2, A_3, . . . A_n, in which the junction A_1A_2 has the temperature, T, while all the remaining junctions have the temperature, T_0, is identical with that in a circuit consisting only of the two metals A_1 and A_2 when the junctions have the temperatures, T and T_0. Finally, we may note the formula :

$$E^{T_2}_{T_1}(AB) + E^{T_3}_{T_2}(AB) + E^{T_4}_{T_3}(AB) + \ldots + E^{T_n}_{T_{n-1}}(AB)$$

$$= E^{T_n}_{T_1}(AB), \quad . \quad . \quad . \quad . \quad . \quad . \quad \textbf{(24·64)}$$

in which $E^{T_4}_{T_3}(AB)$ means the E.M.F. in the circuit AB, the junctions having the temperatures T_3 and T_4 and the positive direction being from A to B at the T_4 junction. The proof may be left to the reader.

§ 24·7. THE APPLICATION OF THERMODYNAMICS TO THERMOELECTRIC PHENOMENA

The electromotive force in a circuit consisting of the two metals A and B in the sense A to B across the junction at the temperature T may be written :

$$E = \Pi - \Pi_0 + K(A) - K_0(A) + K_0(B) - K(B), \quad . \quad \textbf{(24·7)}$$

where Π and K mean the values of these functions corresponding to the temperature, T; while Π_0 and K_0 correspond to the temperature, T_0. Apart from the Joulean development of heat at the rate RC^2, and the transport of heat by conduction, all the phenomena are *reversible* in the thermodynamic sense. If we ignore the latter of these irreversible effects, we can imagine a Carnot engine taking heat from a source at T and rejecting heat to a sink or refrigerator at T_0. We have only to imagine the resistance of the circuit made infinitesimal. The entropy increment due to the passage of the unit quantity of electricity round the circuit will be equal to zero. Therefore

$$0 = \frac{\Pi}{T} - \frac{\Pi_0}{T_0} + \int_{T_0}^{T} \frac{\sigma_A - \sigma_B}{T} dT, \quad . \quad . \quad \textbf{(24·71)}$$

by (24·62). We may regard (24·71) as valid notwithstanding the fact that we have ignored the increment in entropy due to thermal conduction, because this increment is not expressible in electrical terms at all.

If we now imagine the temperature, T_0, maintained constant while T is variable, we get by differentiation of (24·7) and (24·71) :

$$\frac{dE}{dT} = \frac{d\Pi}{dT} + \sigma_A - \sigma_B. \quad . \quad . \quad . \quad (24\cdot72)$$

and

$$\sigma_B - \sigma_A = \frac{d\Pi}{dT} - \frac{\Pi}{T}. \quad . \quad . \quad . \quad (24\cdot73)$$

Elimination of $\sigma_B - \sigma_A$ from the two equations gives us :

$$\Pi = T\frac{dE}{dT}, \quad . \quad . \quad . \quad . \quad (24\cdot74)$$

and, finally, when we substitute this expression for Π in (24·73) we get :

$$\sigma_B - \sigma_A = T\frac{d^2E}{dT^2}. \quad . \quad . \quad . \quad (24\cdot74)$$

The forms of equations (24·73) and (17·63) are identical. The Kelvin coefficient, σ_B, is the exact analogue of the specific heat of a saturated vapour and σ_A that of the saturated liquid. In fact, Kelvin termed the coefficient σ the **specific heat of electricity.** It will be further observed that Π might be described as the **latent heat of evaporation** of electricity out of A into B at the temperature T, and (24·74) is in fact just Clapeyron's equation. The factor corresponding to $V_v - V_l$ of (17·6) is here the quantity of electricity which has crossed the boundary, and this is unity. The phenomena of conduction of electricity in metals and those of thermoelectricity can indeed be fairly adequately accounted for by the simple electron theory (§ 26) in terms of which Π and σ are not merely the analogues of latent heat and specific heat respectively, but are identical with these quantities.

THE ELECTROMAGNETIC THEORY OF CLERK MAXWELL

§ 24·8. ELECTROMAGNETIC WAVES IN AN INSULATING MEDIUM

FOR an isotropic insulating medium the first of the equations (22·72) takes the form :

$$\frac{K}{a}\frac{\partial \mathcal{E}_x}{\partial t} = \frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z}.$$

On differentiating this equation partially with respect to t we obtain

$$\frac{K}{a}\frac{\partial^2 \mathcal{E}_x}{\partial t^2} = \frac{\partial}{\partial y}\left(\frac{\partial H_z}{\partial t}\right) - \frac{\partial}{\partial z}\left(\frac{\partial H_y}{\partial t}\right),$$

and on substituting in this the expressions for $\partial H_y/\partial t$ and $\partial H_z/\partial t$ given in equations (22·73) we get :

$$\frac{\mu K}{a^2}\frac{\partial^2 \mathcal{E}_x}{\partial t^2} = \frac{\partial}{\partial y}\left(\frac{\partial \mathcal{E}_x}{\partial y} - \frac{\partial \mathcal{E}_y}{\partial x}\right) - \frac{\partial}{\partial z}\left(\frac{\partial \mathcal{E}_z}{\partial x} - \frac{\partial \mathcal{E}_x}{\partial z}\right),$$

or
$$\frac{\mu K}{a^2}\frac{\partial^2 \mathcal{E}_x}{\partial t^2} = \left\{\frac{\partial^2 \mathcal{E}_x}{\partial y^2} + \frac{\partial^2 \mathcal{E}_x}{\partial z^2}\right\} - \frac{\partial}{\partial x}\left\{\frac{\partial \mathcal{E}_y}{\partial y} + \frac{\partial \mathcal{E}_z}{\partial z}\right\}.$$

It is obviously desirable to add and subtract the term $\partial^2 \mathcal{E}_x/\partial x^2$ on the right-hand side of the last equation ; so that it may be written :

$$\frac{\mu K}{a^2}\frac{\partial^2 \mathcal{E}_x}{\partial t^2} = \nabla^2 \mathcal{E}_x - \frac{\partial}{\partial x}\,\mathrm{div}\,\mathcal{E}. \qquad . \qquad . \qquad (24\cdot 8)$$

Similarly, by differentiating the first of the equations (22·73), we shall obtain

$$\frac{\mu K}{a^2}\frac{\partial^2 H_x}{\partial t^2} = \nabla^2 H_x - \frac{\partial}{\partial x}\,\mathrm{div}\,\mathbf{H}. \qquad . \qquad . \qquad (24\cdot 81)$$

We have already met with equations of the form of (24·8) and (24·81) in § 9·2 and in the chapter on elasticity. The discussion of equation (9·23) brings out that the *longitudinal*

wave associated with **(24·8)** must travel with *zero* velocity. In other words, there is no such wave. The same thing may be said about equation **(24·81)**. But there is associated with each of these equations a *transverse wave* which travels with the velocity $a/\sqrt{\mu K}\,|$. When the insulating medium is uncharged, **div** $\mathcal{E}$ is zero, while **div** H is zero in any case, and the equations **(24·3)** and **(24·31)** take the familiar form :

$$\left.\begin{array}{l} \dfrac{\mu K}{a^2}\dfrac{\partial^2 \mathcal{E}_x}{\partial t^2} = \nabla^2 \mathcal{E}_x, \\[2mm] \dfrac{\mu K}{a^2}\dfrac{\partial^2 H_x}{\partial t^2} = \nabla^2 H_x. \end{array}\right\} \quad . \quad . \quad . \quad \textbf{(24·811)}$$

The ratio of the velocity of waves in empty space to that in the medium under investigation, i.e. the ratio

$$a/\sqrt{\mu_0 K_0}\,| \,\div\, a/\sqrt{\mu K}\,|,$$

is called the **refractive index** of the medium, and we shall represent it by n. If therefore $\mu_0 = \mu$, as is approximately the case for many media,

$$n = \sqrt{\dfrac{K}{K_0}}\Bigg|. \quad . \quad . \quad . \quad . \quad \textbf{(24·82)}$$

§ 24·9. Electrostatic and Electromagnetic Units

It is well at this stage to recall what has been said about units in § § **22·1** and **24·3**. We have seen that the ratio of the E.M. unit of quantity to the E.S. unit of quantity is numerically equal to c, the number which expresses in *cm. sec.*$^{-1}$ the velocity of electromagnetic waves in empty space. This is the case whether our units are of the ordinary type or the Lorentz-Heaviside type. The experimental measurement of the ratio, and the measurement of the velocity of light waves in free space, both yield, within the limits of experimental error, the same number, differing very little from 3×10^{10}. This result led Maxwell to the view (which cannot now be doubted) that light waves themselves are electromagnetic waves.

§ 25. Relations between the Electric and Magnetic Vectors in Electromagnetic Waves

We shall now study the plane wave described by

$$\left.\begin{array}{l} \mathcal{E}_x = \alpha f(vt - lx - my - nz), \\ \mathcal{E}_y = \beta f(vt - lx - my - nz), \\ \mathcal{E}_z = \gamma f(vt - lx - my - nz), \end{array}\right\} \quad . \quad . \quad \textbf{(25)}$$

in which f is some arbitrarily given function, α, β, γ are the direction cosines of $\mathcal{E}$, the electric intensity, and constants; l, m, n, the direction cosines of the normal to the wave front are also constants, and $v(\equiv a/\sqrt{\mu K})$ is the velocity of propagation. We shall use the abbreviations :

$$\xi \equiv vt - lx - my - nz,$$

$$f'(\xi) \text{ or simply } f' \equiv \frac{df(\xi)}{d\xi},$$

$$f''(\xi) \quad ,, \quad ,, \quad f'' \equiv \frac{d^2f(\xi)}{d\xi^2}.$$

From **(25)** we get

$$\mathbf{div}\ \mathcal{E} = - (\alpha l + \beta m + \gamma n)f'.$$

If the medium be uncharged, **div** $\mathcal{E} = 0$, and consequently

$$\alpha l + \beta m + \gamma n = 0.$$

That is to say, the scalar product of the two unit vectors, (α, β, γ) and (l, m, n), is zero. This means that they are at right angles to one another, or the direction of the electric field intensity (and in an isotropic medium, with which we are now concerned, also that of the electric displacement) is at right angles to the direction of propagation. In this narrower sense therefore the waves are of the transverse type. We have already seen, **§ 24·8**, that they are transverse waves in the wider sense explained in **§ 9·2**, even when **div** $\mathcal{E}$ differs from zero.

Let us now form the **curl** of $\mathcal{E}$, or, say, its X component. We find

$$\frac{\partial \mathcal{E}_z}{\partial y} - \frac{\partial \mathcal{E}_y}{\partial z} = (n\beta - m\gamma)f'.$$

Therefore

$$- \frac{\mu}{a}\frac{\partial H_x}{\partial t} = \frac{(n\beta - m\gamma)}{v}\frac{\partial f}{\partial t},$$

since $\partial f/\partial t = vf'$, or

$$\frac{\partial H_x}{\partial t} = (m\gamma - n\beta)\sqrt{\frac{K}{\mu}}\left|\frac{\partial f}{\partial t}\right..$$

Hence

$$H_x = (m\gamma - n\beta)\sqrt{\frac{K}{\mu}}\left| f + \phi, \right. \qquad . \quad . \quad \textbf{(25·01)}$$

where ϕ is an arbitrary function of x, y, z. We shall ignore ϕ for the moment, and we may write for H_x :

$$H_x = (m\mathcal{E}_z - n\mathcal{E}_y)\sqrt{\frac{K}{\mu}}\ .$$

The expression $(m\mathcal{E}_z - n\mathcal{E}_y)$ is the X component of the vector product of the unit vector (l, m, n), in the direction of propagation of the wave, and the vector $\mathcal{E}$. If we represent the unit vector (l, m, n) by N, we have

$$\mathbf{H} = \sqrt{\frac{K}{\mu}}\ [\mathbf{N}, \mathcal{E}].\qquad .\quad .\quad .\quad .\quad \textbf{(25·02)}$$

In this form the equation shows the relation between the directions of the vectors $\mathcal{E}$, $\mathbf{H}$ and $\mathbf{N}$. The three directions are all at right angles to one another, since, as we have already seen, $\mathbf{N}$ is at right angles to $\mathcal{E}$; and $\mathbf{H}$, being equal to the vector, product of $\mathbf{N}$ and $\mathcal{E}$ multiplied by a scalar quantity, is at right angles to $\mathbf{N}$ and $\mathcal{E}$. Further, if we imagine the direction of $\mathbf{N}$ to rotate towards that of $\mathcal{E}$, **(25·02)** indicates that the direction of $\mathbf{H}$ is that in which an ordinary screw would travel with such a rotation. Or, to put it in another way : if the direction of $\mathcal{E}$ be turned towards that of $\mathbf{H}$, such a rotation would cause a screw to travel in the direction of the propagation of the wave.

Since the absolute value of $\mathbf{N}$ is unity, it follows that the absolute value of $\mathbf{H}$ is :

$$\mathbf{H} = \sqrt{\frac{K}{\mu}}\ \mathcal{E}.\qquad .\quad .\quad .\quad .\quad .\quad \textbf{(25·03)}$$

When the function, f, is a sinusoidal one, it follows that **$\mathbf{H}$ and $\mathcal{E}$ are always in the same phase**, and when K and μ are numerically equal, $\mathbf{H} = \mathcal{E}$.

It follows from **(25·03)** that

$$\frac{\mu \mathbf{H}^2}{8\pi A} = \frac{K\mathcal{E}^2}{8\pi A}\ ;\qquad .\quad .\quad .\quad .\quad .\quad \textbf{(25·04)}$$

so that the electrical energy per unit volume is equal to the magnetic energy per unit volume.

The function, ϕ, which appears in **(25·01)**, *is a function of position only*, and at a given place maintains the same value always. Its physical significance is simply that the wave we have been investigating can pass, without modification, through regions in which permanent fields exist—fields which may be due, for example, to charged bodies, or to permanent magnets.

§ 25·1. Waves in a Conducting Medium

We now turn back to the more general equations **(22·72)**, with the purpose of studying the case where σ differs from zero, i.e. where the medium is not an insulator. The method of **§ 24·8** leads to the equation :

$$\frac{\mu K}{a^2}\frac{\partial^2 \mathcal{E}_x}{\partial t^2} + \frac{4\pi A\mu\sigma}{a^2}\frac{\partial \mathcal{E}_x}{\partial t} = \nabla^2\mathcal{E}_x, \qquad . \quad . \quad \textbf{(25·1)}$$

instead of **(24·8)**, supposing the medium uncharged. Let us write for $\mathcal{E}_x$, adopting Bernoulli's method of solution :

$$\mathcal{E}_x = TR,$$

where T is a function of the time only, while R is a function of x, y and z, and does not involve the time. Substituting in **(25·1)**, and dividing through by the product TR, we get :

$$\frac{\mu K}{a^2}\frac{1}{T}\frac{d^2T}{dt^2} + \frac{4\pi A\mu\sigma}{a^2}\frac{1}{T}\frac{dT}{dt} = \frac{1}{R}\nabla^2 R. \quad . \quad . \quad \textbf{(25·11)}$$

Hence a particular solution of **(25·1)** can be obtained by solving the equations :

$$\left.\begin{aligned}\frac{\mu K}{a^2}\frac{1}{T}\frac{d^2T}{dt^2} + \frac{4\pi A\mu\sigma}{a^2}\frac{1}{T}\frac{dT}{dt} &= m,\\[2mm]\frac{1}{R}\nabla^2 R &= m,\end{aligned}\right\} \quad . \quad . \quad . \quad \textbf{(25·12)}$$

where m is any constant.

We shall try to find a solution representing a plane wave in the direction

$$\mathbf{N} \equiv (l,\, m,\, n),$$

and use ρ in the sense :

$$\rho \equiv lx + my + nz.$$

The T is a function of t only, and we shall select the harmonic function,

$$T = e^{i\omega t}, \quad . \quad . \quad . \quad . \quad . \quad \textbf{(25·13)}$$

where ω is real and positive. On substituting for T in the former of the equations **(25·12)**, we get :

$$-\frac{\mu K}{a^2}\omega^2 + \frac{4\pi A\mu\sigma}{a^2}i\omega = m, \quad . \quad . \quad \textbf{(25·14)}$$

and, in consequence, the latter of the equations **(25·12)** becomes :

$$\frac{d^2R}{d\rho^2} + \left(\frac{\mu K}{a^2}\omega^2 - \frac{4\pi A\mu\sigma}{a^2}i\omega\right)R = 0. \quad . \quad . \quad (25\cdot15)$$

In the case of a perfect insulator, the factor multiplying R is

$$\frac{\mu\omega^2}{a^2}K,$$

whereas when $\sigma \neq 0$ it is

$$\frac{\mu\omega^2}{a^2}K',$$

where

$$K' = K - i\frac{4\pi A\sigma}{\omega}, \quad . \quad . \quad . \quad (25\cdot16)$$

or since

$$\omega = 2\pi/\tau,$$

where τ is the period corresponding to the angular frequency, ω,

$$K' = K - i2A\sigma\tau. \quad . \quad . \quad (25\cdot161)$$

We see therefore that the consequence of the conductivity is that the problem becomes mathematically identical with that of plane waves in a perfect insulator—**provided always that in any plane in which** $\rho(= lx + my + nz)$ **is constant, the field intensity varies in a simple harmonic way.**

We may therefore adopt as particular solutions of **(25·1)** and the corresponding equations associated with the Y and Z axes :

$$\left.\begin{array}{l}\mathcal{E}_x = Le^{i\omega\xi}, \\ \mathcal{E}_y = Me^{i\omega\xi}, \\ \mathcal{E}_z = Ne^{i\omega\xi}, \end{array}\right\} \quad . \quad . \quad . \quad (25\cdot17)$$

where L, M and N are real constants of integration and

$$\xi = t - \frac{\sqrt{\mu K'}}{a}\bigg| (lx + my + nz). \quad . \quad . \quad (25\cdot171)$$

It is clear that equations **(25·17)** may be written :

$$\left.\begin{array}{l}\mathcal{E}_x = \alpha\mathcal{E}e^{i\omega\xi}, \\ \mathcal{E}_y = \beta\mathcal{E}e^{i\omega\xi}, \\ \mathcal{E}_z = \gamma\mathcal{E}e^{i\omega\xi}, \end{array}\right\} \quad . \quad . \quad . \quad (25\cdot172)$$

(α, β, γ) representing the direction of $\mathcal{E}$. On forming the **divergence** of $\mathcal{E}$ from **(25·172)** we find again, as we found for an insulator in **§ 25**,

$$\alpha l + \beta m + \gamma n = 0$$

and the electrical field intensity is orthogonal to the direction, $\mathbf{N} \equiv (l, m, n)$, along which, as we shall find, a wave is being propagated.

We are now able to turn our co-ordinate axes so that the X axis coincides with $\boldsymbol{\mathcal{E}}$, and one of the others, say the Z axis, coincides with $\mathbf{N}$. Equations (25·172) now become simply :

$$\mathcal{E} = \mathcal{E}e^{i\omega\left(t - \frac{\sqrt{\mu K'}|}{a}z\right)}, \quad . \quad . \quad . \quad . \quad (25\text{·}173)$$

where $\mathcal{E}$ is the electric field intensity and $\mathcal{E}$ its amplitude.

As K' is complex, we shall write provisionally,

$$\frac{\sqrt{\mu K'}}{a} \equiv \alpha + i\beta. \quad . \quad . \quad . \quad (25\text{·}174)$$

in which α and β are *real* constants. Equation (25·173) then gives us :

$$\mathcal{E} = \mathcal{E}e^{\omega\beta z}e^{i\omega(t - \alpha z)}. \quad . \quad . \quad . \quad (25\text{·}18)$$

The constant α is positive, and if v be the velocity of propagation of the plane wave,

$$\alpha = 1/v. \quad . \quad . \quad . \quad . \quad (25\text{·}181)$$

It will be observed, further, that β is a negative constant, and it is convenient to write it in the form :

$$\beta = -\frac{h}{v}, \quad . \quad . \quad . \quad . \quad (25\text{·}182)$$

h being a *positive* constant. Hence

$$\mathcal{E} = \mathcal{E}e^{-\frac{\omega h z}{v}}e^{i\omega\left(t - \frac{z}{v}\right)},$$

or since $\qquad\qquad \omega = 2\pi/\tau$

and $\qquad\qquad\qquad v\tau = \lambda,$

where λ is the wave-length,

$$\mathcal{E} = \mathcal{E}e^{-\frac{2\pi h z}{\lambda}}e^{i2\pi\left(\frac{t}{\tau} - \frac{z}{\lambda}\right)}. \quad . \quad . \quad . \quad (25\text{·}183)$$

The following relations can easily be inferred from the definitions of α, β, v and h :

$$\left.\begin{aligned} \frac{\mu K}{a^2} &= \frac{1 - h^2}{v^2}, \\ \frac{A\sigma\mu\tau}{a^2} &= \frac{h}{v^2}; \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad (25\text{·}19)$$

and by division :

$$\frac{K}{A\sigma\tau} = \frac{1 - h^2}{h}. \quad . \quad . \quad . \quad (25\text{·}191)$$

If c be the velocity of waves in empty space, K_0 the dielectric constant of empty space, and if the μ of the conducting medium be practically identical with that of empty space, then

$$c^2 = \frac{a^2}{\mu K_0},$$

or $$\mu/a^2 = 1/K_0 c^2,$$

and therefore equations **(25·19)** may be written, if we replace c/v by n, the **refractive index** of the conducting medium (**§ 24·8**),

$$\left. \begin{aligned} \frac{K}{K_0} &= n^2(1 - h^2), \\ \frac{A\sigma\tau}{K_0} &= n^2 h. \end{aligned} \right\} \qquad . \quad . \quad . \quad \textbf{(25·192)}$$

§ 25·2. Critical Observations on Maxwell's Theory

The results of **§ 24·8** lead actually or apparently to a unique (phase) velocity of propagation of electromagnetic waves in an insulating medium. They therefore require some qualification or amplification before they can become a satisfactory basis for the theory of light. Two important assumptions were made in arriving at them : namely that μ and K are constants, the values of which depend on the medium and on nothing else. So far as μ is concerned this is practically true for many media, and we may let the assumptions concerning it stand. The corresponding assumptions concerning K are much more serious. They are intimately associated with the tacit assumption in our theory of the homogeneity, or rather, of the continuous character of the medium. Now K, it will be observed, enters Maxwell's equations through the formula :

$$\mathbf{D} = \frac{K}{4\pi A}\mathcal{E}, \qquad . \quad . \quad . \quad . \quad \textbf{(25·2)}$$

which in fact defines K for the purposes of Maxwell's theory. All material media contain, or are constituted of minute charged particles, the displacement of which under the field intensity, $\mathcal{E}$, contributes to **D**. We shall learn in **§ 26·4** that when $\mathcal{E}$ varies in a simple harmonic way K is *a function, not only of the medium, but of the frequency*, ω. For a given medium it has a definite value which is constant (i.e. independent of $\mathcal{E}$), provided time variations are simple harmonic variations of some definite frequency, ω. The theory which we shall develop in **§§ 26·2**

et seq. is in fairly satisfactory accord with the observed dependence of the velocity of light on the frequency in many media, and it makes K approach asymptotically to a limiting value, agreeing with the K of electrostatics, as ω approaches the limit zero. Maxwell's theory itself is not at fault nor are the equations **(22·72)** and **(22·73)** provided we apply them to simple harmonic waves and interpret K correctly.

Similar remarks may be made about propagation in conducting media. Experiment supports the theory of §§ **24·8– 25·1** for *long waves*, i.e. for *small values of* ω, in which case K and σ are near the limiting values corresponding to $\omega = 0$, and sensibly independent of ω. We shall return to this subject in § **25·9.**

§ 25·3. Poynting's Theorem

Writing the equations **(22·72)** and **(22·73)** in the form :

$$\frac{K}{a}\frac{\partial\mathcal{E}}{\partial t} + \frac{4\pi A\sigma}{a}\mathcal{E} = \text{curl } \mathbf{H},$$

$$-\frac{\mu}{a}\frac{\partial\mathbf{H}}{\partial t} = \text{curl } \mathcal{E},$$

and forming the scalar products with $\mathcal{E}$ and $\mathbf{H}$ respectively, we find, on subtracting :

$$\frac{K}{a}\left(\mathcal{E}, \frac{\partial\mathcal{E}}{\partial t}\right) + \frac{\mu}{a}\left(\mathbf{H}, \frac{\partial\mathbf{H}}{\partial t}\right) + \frac{4\pi A\sigma}{a}\mathcal{E}^2 = (\mathcal{E}, \text{curl } \mathbf{H}) - (\mathbf{H}, \text{curl } \mathcal{E}),$$

and therefore, by **(2·44)**,

$$\frac{\partial}{\partial t}\left\{\frac{K}{2a}\mathcal{E}^2 + \frac{\mu}{2a}\mathbf{H}^2\right\} + \frac{4\pi A}{a}\sigma\mathcal{E}^2 = \text{div } [\mathbf{H}, \mathcal{E}]. \quad . \quad (25\cdot3)$$

On multiplying both sides by $a/4\pi A$, we get :

$$\frac{\partial}{\partial t}\left\{\frac{K}{8\pi A}\mathcal{E}^2 + \frac{\mu}{8\pi A}\mathbf{H}^2\right\} + \sigma\mathcal{E}^2 = \frac{a}{4\pi A}\text{div } [\mathbf{H}, \mathcal{E}].$$

The quantity within the brackets { } is the electromagnetic energy per unit volume, which we shall denote by U, while $\sigma\mathcal{E}^2$ represents the rate, per unit volume, of conversion of electromagnetic energy into heat. This follows from **(21·93)**, since the conduction current density is

$$\mathbf{i} = \sigma\mathcal{E}.$$

Therefore

$$-\frac{\partial U}{\partial t} = \sigma \mathcal{E}^2 + \frac{a}{4\pi A} \, \text{div} \, [\mathcal{E}, \mathbf{H}], \quad . \quad . \quad (25\cdot31)$$

in which we have effected the change of sign in the vector product by inverting the order of $\mathcal{E}$ and $\mathbf{H}$. We now multiply (25·31) by the volume element, $dx \, dy \, dz$, and integrate over the whole of the volume within some closed surface. This gives us :

$$-\frac{\partial}{\partial t} \begin{Bmatrix} \text{Electromagnetic energy} \\ \text{within closed surface} \end{Bmatrix} = \begin{Bmatrix} \text{Rate of production of heat} \\ \text{within closed surface} \end{Bmatrix}$$

$$+ \frac{a}{4\pi A} \iint ([\mathcal{E}, \mathbf{H}], \, d\mathbf{S}), \quad . \quad . \quad . \quad (25\cdot32)$$

in which the volume integration

$$\iiint \text{div} \, [\mathcal{E}, \mathbf{H}] \, dx \, dy \, dz$$

has been expressed appropriately as a surface integral by the theorem of Gauss (3·01). The statement (25·32) expresses **Poynting's Theorem.** It asserts that the rate of diminution of the electromagnetic energy within a region enclosed by a surface is equal to the rate of generation of heat within the region plus an integral extended over the surface. It is obvious that this integral expresses the rate at which electromagnetic energy flows outwards through the surface. Consequently the vector,

$$\mathbf{p} \equiv \frac{a}{4\pi A}[\mathcal{E}, \mathbf{H}], \quad . \quad . \quad . \quad (25\cdot33)$$

apparently [1] represents in amount, and in direction, the density of the flow of electromagnetic energy, i.e. the quantity passing per unit time per unit area normal to the direction of flow. The vector, **p**, is known as **Poynting's vector.**

§ 25·4. ELECTROMAGNETIC WAVES IN ANISOTROPIC MEDIA

In dealing with the propagation of electromagnetic waves in anisotropic media, we shall confine our attention, as in fact we have always done so far, to the case where μ, the permeability of the medium, is a scalar quantity, constant (i.e. independent

---

[1] This qualification is necessary since Poynting's theorem is not affected by superposing on **p** any vector the divergence of which is zero. The density of the flow of electromagnetic energy ought therefore to be expressed in the form **p** + **p′**, where **div p′** = 0.

of **H**) and not appreciably different from the permeability of empty space. We assume therefore that the anisotropic character of the medium finds its expression in the tensor character of K. Reference to **(19·23)** will show that equations **(22·72)**, if the medium is an insulating one ($\sigma = 0$), must take the form of :

$$\frac{1}{a}\left\{K_{xx}\frac{\partial\mathcal{E}_x}{\partial t} + K_{xy}\frac{\partial\mathcal{E}_y}{\partial t} + K_{xz}\frac{\partial\mathcal{E}_z}{\partial t}\right\} = \{\mathbf{curl\ H}\}_x, \quad . \quad (25\cdot4)$$

and two further equations similarly associated with the Y and Z axes respectively. If the axes of co-ordinates be made to coincide with the principal axes of displacement, the equations **(25·4)** simplify, as shown in **§ 19·3**, to :

$$\left.\begin{aligned}
\frac{K_1}{a}\frac{\partial\mathcal{E}_x}{\partial t} &= \frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z}, \\
\frac{K_2}{a}\frac{\partial\mathcal{E}_y}{\partial t} &= \frac{\partial H_x}{\partial z} - \frac{\partial H_z}{\partial x}, \\
\frac{K_3}{a}\frac{\partial\mathcal{E}_z}{\partial t} &= \frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y},
\end{aligned}\right\} \quad . \quad . \quad . \quad (25\cdot41)$$

in which the principal dielectric constants, K_{xx}, K_{yy} and K_{zz} are represented by K_1, K_2 and K_3 respectively for brevity. The divergence equation,

$$\mathbf{div\ D} = 0, \quad . \quad . \quad . \quad . \quad (25\cdot411)$$

(we are supposing the medium to be uncharged) may be written :

$$K_1\frac{\partial\mathcal{E}_x}{\partial x} + K_2\frac{\partial\mathcal{E}_y}{\partial y} + K_3\frac{\partial\mathcal{E}_z}{\partial z} = 0. \quad . \quad . \quad (25\cdot412)$$

Applying the method of **§ 24·8** to the equations **(25·41)** we get :

$$\frac{\mu K_1}{a^2}\frac{\partial^2\mathcal{E}_x}{\partial t} = \nabla^2\mathcal{E}_x - \frac{\partial}{\partial x}\mathbf{div\ \mathcal{E}}, \quad . \quad . \quad (25\cdot42)$$

and similar equations associated with the Y and Z axes ; but we may not now, in general, proceed further by equating **div ε** to zero. If however we attempt, in the first instance, to find a particular solution for which $\mathcal{E}_x = \mathcal{E}$, $\mathcal{E}_y = 0$, $\mathcal{E}_z = 0$, we shall then have from **(25·412)**

$$K_1\frac{\partial\mathcal{E}_x}{\partial x} = 0,$$

and hence $\qquad\qquad \mathbf{div\ \mathcal{E}} = \dfrac{\partial\mathcal{E}_x}{dx} = 0,$

so that in this special case **(25·42)** becomes :

$$\frac{\mu K_1}{a^2} \frac{\partial^2 \mathcal{E}_x}{\partial t^2} = \nabla^2 \mathcal{E}_x, \qquad . \quad . \quad . \quad . \quad \textbf{(25·43)}$$

of which $\qquad \mathcal{E}_x = f_x(vt - lx - my - nz)$

will be a solution representing a plane wave, if

$$v = \frac{a}{\sqrt{\mu K_1 |}}. \qquad . \quad . \quad . \quad . \quad \textbf{(25·44)}$$

Since $\mathcal{E}_y = 0$ and $\mathcal{E}_z = 0$, we have

$$\textbf{div } \mathcal{E} = \frac{\partial \mathcal{E}_x}{\partial x} = - l f_x' = 0,$$

where $\qquad f_x' \equiv \dfrac{d f_x(\xi)}{d\xi},$

and $\qquad \xi = vt - lx - my - nz.$

Hence we must infer

$$l = 0 \ ;$$

or the possible directions of propagation of this plane wave are, all of them, at right angles to the X principal axis.

We shall represent the velocity **(25·44)**, which we shall term a principal velocity, by **a**, so that

$$\textbf{a} = a/\sqrt{\mu K_1}. \qquad . \quad . \quad . \quad \textbf{(25·441)}$$

Similarly, we may derive solutions of the differential equations **(25·42)** associated with the Y and Z principal axes, for which the velocities are

$$\left. \begin{array}{l} \textbf{b} = a/\sqrt{\mu K_2 |}, \\[2mm] \textbf{c} = a/\sqrt{\mu K_3 |}. \end{array} \right\} \qquad . \quad . \quad . \quad \textbf{(25·442)}$$

It is convenient to suppose the co-ordinate axes so placed that, when **a, b** and **c** are all different,

$$\textbf{a} > \textbf{b} > \textbf{c}. \qquad . \quad . \quad . \quad . \quad \textbf{(25·443)}$$

We now proceed to find a solution of equations **(25·42)** which will represent a plane wave propagated in any direction (l, m, n). To do this we try

$$\left. \begin{array}{l} D_x = \alpha f(\xi), \\[1mm] D_y = \beta f(\xi), \\[1mm] D_z = \gamma f(\xi), \end{array} \right\} \qquad . \quad . \quad . \quad . \quad \textbf{(25·45)}$$

in which $\textbf{D} \equiv (D_x, D_y, D_z)$ is the Maxwell displacement, α, β, γ are its direction cosines, and ξ has the meaning already explained.

Since D_x differs from $K_1 \mathcal{E}_x$ only by the constant $1/4\pi A$, it will be simpler to write, instead of **(25·45)**,

$$
\left.
\begin{aligned}
\mathcal{E}_x &= \frac{\alpha}{K_1} f(\xi), \\[4pt]
\mathcal{E}_y &= \frac{\beta}{K_2} f(\xi), \\[4pt]
\mathcal{E}_z &= \frac{\gamma}{K_3} f(\xi).
\end{aligned}
\right\}
\qquad \textbf{(25·451)}
$$

We easily find :

$$
\left.
\begin{aligned}
\frac{\partial^2 \mathcal{E}_x}{\partial t^2} &= \frac{\alpha}{K_1} v^2 f'', \\[6pt]
\operatorname{\mathbf{div}} \mathcal{E} &= -\left\{ \frac{\alpha l}{K_1} + \frac{\beta m}{K_2} + \frac{\gamma n}{K_3} \right\} f', \\[6pt]
\frac{\partial}{\partial x}(\operatorname{div} \mathcal{E}) &= l\frac{\alpha l}{K_1}\left\{ + \frac{\beta m}{K_2} + \frac{\gamma n}{K_3} \right\} f'', \\[6pt]
\nabla^2 \mathcal{E}_x &= \frac{\alpha}{K_1} f'',
\end{aligned}
\right\}
\qquad \textbf{(25·46)}
$$

and

remembering that $l^2 + m^2 + n^2 = 1$.

On substituting these expressions in **(25·42)** we find :

$$
\frac{\mu \alpha}{a^2} v^2 = \frac{\alpha}{K_2} - lG,
$$

where

$$
G \equiv \frac{\alpha l}{K_1} + \frac{\beta m}{K_2} + \frac{\gamma n}{K_3}. \qquad \textbf{(25·461)}
$$

This result may be written in the form :

$$
\alpha v^2 = \alpha \mathbf{a}^2 - lG',
$$

where

$$
G' \equiv \frac{a^2 G}{\mu} \equiv \mathbf{a}^2 \alpha l + \mathbf{b}^2 \beta m + \mathbf{c}^2 \gamma n, \qquad \textbf{(25·462)}
$$

by **(25·441)** and **(25·461)**, and there are of course two further equations similarly associated with the Y and Z axes. We may write the three equations in the following way :

$$
\left.
\begin{aligned}
\frac{l}{\mathbf{a}^2 - v^2} &= \frac{\alpha}{G'}, \\[6pt]
\frac{m}{\mathbf{b}^2 - v^2} &= \frac{\beta}{G'}, \\[6pt]
\frac{n}{\mathbf{c}^2 - v^2} &= \frac{\gamma}{G'}.
\end{aligned}
\right\}
\qquad \textbf{(25·47)}
$$

On multiplying these three equations by l, m, and n respectively and adding, we obtain **Fresnel's equation** for the velocity of a plane wave in a given direction (l, m, n), namely :

$$\frac{l^2}{v^2 - a^2} + \frac{m^2}{v^2 - b^2} + \frac{n^2}{v^2 - c^2} = 0, \qquad . \quad . \quad (25\cdot48)$$

since $\qquad\qquad \alpha l + \beta m + \gamma n = 0,$

as we can prove by substituting equations **(25·451)** in **(25·411)**. Since (α, β, γ) represents the direction of the *electric displacement* in the wave, the last equation asserts that **the direction of the electric displacement is at right angles to the direction of propagation.** The electric field intensity is therefore not in general at right angles to the direction of propagation, since it does not in general coincide with that of the displacement **(§ 19·3).** The consequences of Fresnel's equation will be discussed in the chapter devoted to the propagation of light in crystalline media.

§ 25·5. REFLEXION OF ELECTROMAGNETIC WAVES AT THE INTERFACE BETWEEN TWO ISOTROPIC MEDIA—ELECTRIC FIELD PERPENDICULAR TO PLANE OF INCIDENCE

Imagine a plane wave incident at an angle, θ, at the plane surface separating two dielectric media. The angle, θ, is the angle between a normal to a wave front, i.e. to a plane of constant phase, and the normal to the interfacial boundary where the former meets it. These two normals define a plane—the **plane of incidence**—and we shall consider in the first place the case where the plane wave is polarized with its electric vibrations perpendicular to the plane of incidence. In Fig. 25·5 SS represents the boundary surface (perpendicular to the plane of the paper) between the two media, (a) and (b), and AO represents the direction of propagation of the plane wave. We may conveniently describe the wave by reference to rectangular coordinates, the Z axis being in the direction of propagation, the X axis in the plane of incidence (coincident with the plane of the paper) and the Y axis directed upwards from the plane of the paper. The equation of the wave may be represented by

$$\left.\begin{array}{l} \mathcal{E}_x = 0, \\[2mm] \mathcal{E}_y = \mathcal{E}_i \cos \omega \left(t - \dfrac{z}{v}\right), \\[2mm] \mathcal{E}_z = 0, \end{array}\right\} \qquad . \quad . \quad . \quad (25\cdot5)$$

where $\mathcal{E}_i$ represents the amplitude of the electric field intensity

in the wave and $v = a/\sqrt{\mu K_a}\,$. K_a is the dielectric constant of the medium a, while μ is its permeability. We shall assume μ to have the same value for both of the media a and b, and to
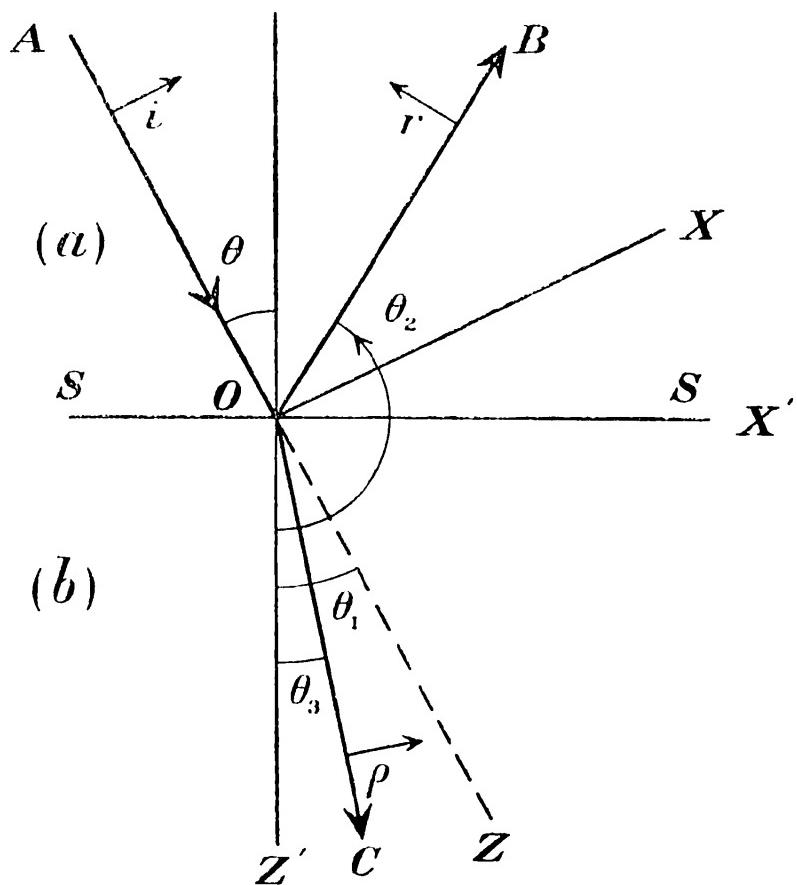
Fig. 25·5

be independent of the magnetic field intensity. This is practically the case for many dielectric media. It is preferable to use instead of equations (25·5) the complex expressions :

$$
\left.
\begin{aligned}
\mathcal{E}_x &= 0, \\
\mathcal{E}_y &= \mathcal{E}_i e^{i\omega(t - \frac{z}{v})}, \\
\mathcal{E}_z &= 0.
\end{aligned}
\right\} \qquad \cdots \quad \textbf{(25·51)}
$$

The use of complex expressions may be justified in much the same way as in § (23·7). The corresponding equations for the magnetic field intensity in the wave are, by (25·01),

$$
\left.
\begin{aligned}
H_x &= -\sqrt{\frac{K_a}{\mu}}\; \mathcal{E}_i e^{i\omega(t - \frac{z}{v})}, \\
H_y &= 0, \\
H_z &= 0.
\end{aligned}
\right\} \qquad \cdots \quad \textbf{(25·52)}
$$

Equations (25·51) and (25·52) describe the wave completely. To find out what happens at the boundary we have at our disposal the boundary conditions expressed by equations (20·32) and (20·34). The only way of satisfying these conditions is to assume the existence of a plane reflected wave travelling back into medium a, and a plane refracted wave travelling onwards into medium b. Hence the law that the reflected (or refracted) ray is in the plane of incidence (cf. § 28·5). The axes of coordinates we have been using, while convenient for the description of the incident wave, are unsuitable for the further investigation, and we shall now introduce new axes so placed that one of them, the Z axis, is directed from a to b perpendicular to the interfacial plane. We shall leave the direction of the Y axis unchanged. We may distinguish the new axes by dashes, thus : X', Y', Z'. The scheme of transformation from X, Y Z, to X', Y', Z', in accordance with § 2·2, is described in the subjoined table :

	X	Y	Z
X'	$\cos \theta$	0	$\sin \theta$
Y'	0	1	0
Z'	$-\sin \theta$	0	$\cos \theta$

Referred to the new axes, equations (25·51) and (25·52) become :

$$\left. \begin{aligned} \mathcal{E}_x' &= 0, \\ \mathcal{E}_y' &= \mathcal{E}_i e^{i\omega\left(t - \frac{\sin \theta . x' + \cos \theta . z'}{v}\right)}, \\ \mathcal{E}_z' &= 0, \end{aligned} \right\} \qquad . \quad . \quad (25\cdot53)$$

$$\left. \begin{aligned} H_x' &= -\cos \theta \sqrt{\frac{K_a}{\mu}} \; \mathcal{E}_i e^{i\omega\left(t - \frac{\sin \theta . x' + \cos \theta z'}{v}\right)}, \\ H_y' &= 0, \\ H_z' &= \sin \theta \sqrt{\frac{K_a}{\mu}} \; \mathcal{E}_i e^{i\omega\left(t - \frac{\sin \theta . x' + \cos \theta . z'}{v}\right)}. \end{aligned} \right\} . \quad (25\cdot54)$$

The angle, θ, it may be observed, is the angle between the direction of the wave and the direction of the Z' axis. In what follows we may drop the dashes, since it will not be necessary to make any further use of the original axes. We shall have another set of equations similar to (25·53) and (25·54) describing the reflected wave, and still another set describing the refracted

wave. The reader will be able to write down all three sets for himself. The amplitudes may be distinguished by :

$$\mathcal{E}_i\text{—incident wave,}$$

$$\mathcal{E}_r\text{—reflected wave,}$$

$$\mathcal{E}_\rho\text{—refracted wave,}$$

while the corresponding angles with the new Z axis may be distinguished one from another by writing

$$\left.\begin{array}{l} \theta_1 = \theta,\ \text{angle of incidence,} \\ \theta_2 = \pi - \phi,\ \text{where } \phi \text{ is the angle of reflexion,} \\ \theta_3 = \psi,\ \text{angle of refraction.} \end{array}\right\} \quad \textbf{(25·541)}$$

In the reflected wave we shall have the dielectric constant K_a, and in the refracted one, K_b.

Applying the boundary condition **(20·32)** to the electric field intensities, we get :

$$\mathcal{E}_i e^{i\omega\left(t-\frac{\sin\theta_1.x}{v_a}\right)} + \mathcal{E}_r e^{i\omega\left(t-\frac{\sin\theta_2.x}{v_a}\right)} = \mathcal{E}_\rho e^{i\omega\left(t-\frac{\sin\theta_3.x}{v_b}\right)}. \quad . \quad \textbf{(25·55)}$$

We have put $z = 0$, since we are dealing with conditions at the boundary. The equation **(25·55)** must hold at all times and for all values of x. Putting t and x equal to zero, we get :

$$\mathcal{E}_i + \mathcal{E}_r = \mathcal{E}_\rho. \quad . \quad . \quad . \quad \textbf{(25·551)}$$

It is now clear that all three exponentials in **(25·55)** are equal, and consequently

$$\frac{\sin\theta_1.x}{v_a} = \frac{\sin\theta_2.x}{v_a} = \frac{\sin\theta_3.x}{v_b}.$$

Therefore

$$\sin\theta_1 = \sin\theta_2,$$

or

$$\sin\theta = \sin\phi,$$

and

$$\theta = \phi,$$

which is the familiar law of reflexion. Further,

$$\frac{\sin\theta_1}{v_a} = \frac{\sin\theta_3}{v_b},$$

or

$$\frac{v_a}{v_b} = \frac{\sqrt{K_b\,|}}{\sqrt{K_a\,|}} = \frac{\sin\theta_1}{\sin\theta_3} = \frac{\sin\theta}{\sin\psi}, \quad . \quad \textbf{(25·552)}$$

the familiar law of refraction. The boundary condition **(20·32)**

applied the component of the magnetic field intensity parallel to the boundary gives us :

$$- \cos \theta_1 \sqrt{\frac{K_a}{\mu}} \; \mathcal{E}_i - \cos \theta_2 \sqrt{\frac{K_a}{\mu}} \; \mathcal{E}_r$$

$$= - \cos \theta_3 \sqrt{\frac{K_b}{\mu}} \; \mathcal{E}_\rho, \qquad . \quad . \quad . \quad . \quad \textbf{(25·56)}$$

while

$$\cos \theta_1 = \cos \theta,$$

$$\cos \theta_2 = - \cos \phi = - \cos \theta,$$

$$\cos \theta_3 = \cos \psi,$$

and

$$\sqrt{\frac{K_b}{K_a}} = \frac{\sin \theta}{\sin \psi}.$$

Therefore

$$\mathcal{E}_i - \mathcal{E}_r = \frac{\cos \psi \sin \theta}{\cos \theta \sin \psi} \mathcal{E}_\rho. \quad . \quad . \quad . \quad \textbf{(25·561)}$$

Adding and subtracting equations **(25·551)** and **(25·561)** and dividing one result by the other, we get :

$$\frac{\mathcal{E}_r}{\mathcal{E}_i} = \frac{1 - \dfrac{\cos \psi \sin \theta}{\cos \theta \sin \psi}}{1 + \dfrac{\cos \psi \sin \theta}{\cos \theta \sin \psi}},$$

or

$$\frac{\mathcal{E}_r}{\mathcal{E}_i} = \frac{\sin \psi \cos \theta - \cos \psi \sin \theta}{\sin \psi \cos \theta + \cos \psi \sin \theta},$$

and consequently

$$\mathcal{E}_r = \mathcal{E}_i \frac{\sin (\psi - \theta)}{\sin (\psi + \theta)}. \qquad . \quad . \quad . \quad . \quad \textbf{(25·57)}$$

As we approach perpendicular incidence **(25·57)** becomes

$$\mathcal{E}_r = \mathcal{E}_i \frac{1 - \dfrac{\theta}{\psi}}{1 + \dfrac{\theta}{\psi}}.$$

When therefore the refractive index from a to b is *greater* than unity $\mathcal{E}_r$ will be *negative*. It is usual to keep the amplitude positive, and in order to do this in the present case it is necessary

to introduce a phase change of π. Such a change in phase will not occur when the refractive index from a to b is *less* than unity.

Substituting the expression (25·57) in (25·551) we find for the amplitude of the refracted wave :

$$\mathcal{E}_\rho = \frac{2 \sin \psi \cos \theta}{\sin (\psi + \theta)} \mathcal{E}_i. \qquad . \quad . \quad . \quad . \quad \textbf{(25·58)}$$

The amplitudes of the magnetic field intensities in the reflected and refracted waves are also given by (25·57) and (25·58), since H is proportional to $\mathcal{E}$. It is usual to employ the square of the electric amplitude as a measure of the so-called **intensity** in the wave. This term is rather unfortunate, since it is also used in a somewhat different sense in the description of the fields of force in the region occupied by the wave. The ideal expression for the intensity of the wave would be Poynting's vector (25·33) which is identical with the product of the phase velocity and the energy per unit volume. So that, if I represents the intensity,

$$I = \frac{a}{4\pi A} \sqrt{\frac{K}{\mu}} \, \mathcal{E}^2, \quad . \quad . \quad . \quad \textbf{(25·59)}$$

where a and A are the two units constants.

§ 25·6. ELECTRIC FIELD PARALLEL TO PLANE OF INCIDENCE

We now turn to the case where the electric vibrations are parallel to the plane of incidence and follow the procedure of **§ 25·5**. The amplitudes in the three waves are directed as indicated by the arrows i, r and ρ, in Fig 25·5, and the description of the incident wave, when referred to the original axes, will be :

$$\left.\begin{array}{l} \mathcal{E}_x = \mathcal{E}_i e^{i\omega\left(t-\frac{z}{v}\right)}, \\[4pt] \mathcal{E}_y = 0, \\[4pt] \mathcal{E}_z = 0, \end{array}\right\} \qquad . \quad . \quad . \quad . \quad \textbf{(25·6)}$$

$$\left.\begin{array}{l} H_x = 0, \\[4pt] \mathbf{H}_y = \sqrt{\frac{K}{\mu}} \, \mathcal{E}_i \, e^{i\omega\left(t-\frac{z}{v}\right)}, \\[4pt] H_z = 0. \end{array}\right\} \qquad . \quad . \quad \textbf{(25·61)}$$

We now turn the axes (as in the last section) about the Y axis, so as to cause the Z axis to point vertically downwards, i.e. from a to b, and work out expressions corresponding to (25·53) and (25·54) by using the table of direction cosines as in § 25·5. For the purpose of finding the relationships between the

amplitudes of the incident, reflected and refracted waves, we need only the expressions for the electric field intensities. The two boundary conditions, $(20\cdot32)$ and $(20\cdot34)$, yield respectively :

$$\left.\begin{aligned}\mathcal{E}_i \cos\theta_1 + \mathcal{E}_r \cos\theta_2 &= \mathcal{E}_\rho \cos\theta_3, \\ K_a\mathcal{E}_i \sin\theta_1 + K_a\mathcal{E}_r \sin\theta_2 &= K_b\mathcal{E}_\rho \sin\theta_3. \end{aligned}\right\} \quad . \quad (25\cdot62)$$

These equations are seen to be equivalent, when we remember equations $(25\cdot541)$ and $(25\cdot552)$, to

$$\left.\begin{aligned}\mathcal{E}_i - \mathcal{E}_r &= \frac{\cos\psi}{\cos\theta}\mathcal{E}_\rho, \\ \mathcal{E}_i + \mathcal{E}_r &= \frac{\sin\theta}{\sin\psi}\mathcal{E}_\rho, \end{aligned}\right\} \quad . \quad . \quad . \quad (25\cdot63)$$

from which we infer :

$$\mathcal{E}_r = \mathcal{E}_i\frac{\tan(\theta - \psi)}{\tan(\theta + \psi)}, \quad . \quad . \quad . \quad . \quad (25\cdot64)$$

$$\mathcal{E}_\rho = \mathcal{E}_i\frac{2\sin\psi\cos\theta}{\sin(\theta + \psi)\cos(\theta - \psi)}. \quad . \quad . \quad (25\cdot65)$$

§ 25·7. Polarization by Reflexion—Brewster's Law

When the direction of vibration in the incident wave makes any angle, χ, with the plane of incidence, we may resolve the field vectors, $\mathcal{E}$ and H, into their components perpendicular to, and in the plane of incidence, and the results of § 25·5 and § 25·6 respectively give us complete information about the corresponding amplitudes in the reflected and refracted waves. It is obvious that there is a value of the angle θ (or the corresponding ψ) such that $\theta + \psi$ is a right angle. For this particular pair of values $\tan(\theta + \psi)$ is infinite, and $(25\cdot64)$ leads to the conclusion that in this case the reflected wave can only have electric vibrations perpendicular to the plane of incidence. It is completely (plane) polarized. The plane of incidence is called the **plane of polarization** of the reflected waves. Each of the particular values of θ and ψ, which we may write θ_p and ψ_p, is called an **angle of polarization**. Since $\theta_p + \psi_p = \pi/2$, it follows that $\cos\theta_p = \sin\psi_p$ and $\sin\theta_p = \cos\psi_p$. Consequently the refractive index from a to b obeys the law :

$$\left.\begin{aligned}{}_an_b &= \tan\theta_p, \\ {}_bn_a &= \tan\psi_p. \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad (25\cdot7)$$

This is **Brewster's Law.**

§ 25·8. Reflexion of Electromagnetic Waves at the Interface between an Insulating and a Conducting Medium

We learned in § 25·1 that problems concerning harmonic waves in conducting media can be made mathematically identical with the corresponding problems for insulating media. For simplicity we shall confine our attention to perpendicular incidence, and we may suppose the medium a to be evacuated space and b to be the conducting medium. For insulating media and perpendicular incidence we have

$$\mathcal{E}_r = \mathcal{E}_i \frac{1 - 0/\psi}{1 + 0/\psi},$$

the ratio $0/\psi$ being the limiting value of θ/ψ for $\theta = 0$. Therefore with perpendicular incidence

$$\mathcal{E}_r = \mathcal{E}_i \frac{\sqrt{K_b} \mid - \sqrt{K_a} \mid}{\sqrt{K_b} \mid + \sqrt{K_a} \mid}.$$

It follows, if K' is the complex quantity described in § 25·1, and if K_0 is the value of K for empty space,

$$\mathcal{E}_r = \mathcal{E}_i \frac{\sqrt{K'} \mid - \sqrt{K_0} \mid}{\sqrt{K'} \mid + \sqrt{K_0} \mid}, \quad \cdot \quad \cdot \quad \cdot \quad (25\cdot8)$$

where $\mathcal{E}_r$ is the amplitude (possibly complex) of the reflected wave. Now

$$\sqrt{\mu K'} \mid /a = \frac{1}{v} - i\frac{h}{v},$$

(§ 25·1) and since $\sqrt{\mu K_0}/a^2 = 1/c$, where c is the velocity of electromagnetic waves in empty space,

$$\sqrt{K'/K_0} \mid = n - i \, nh,$$

where n is the (real) refractive index of the conducting medium. Hence (25·8) becomes :

$$\mathcal{E}_r = \mathcal{E}_i \frac{n - 1 - i \, nh}{n + 1 - i \, nh}. \quad \cdot \quad \cdot \quad \cdot \quad (25\cdot81)$$

Therefore

$$\mathcal{E}_r = \frac{\sqrt{(n-1)^2 + n^2 h^2} \mid}{\sqrt{(n+1)^2 + n^2 h^2} \mid} \mathcal{E}_i e^{i(\phi_1 - \phi_2)},$$

where

$$\phi_1 = \tan^{-1}(-nh/(n-1)) \quad \text{and} \quad \phi_2 = \tan^{-1}(-nh/(n+1)).$$

[1] We are assuming that both media have the same permeability.

Therefore

$$\begin{Bmatrix} \text{Amplitude of} \\ \text{reflected wave} \end{Bmatrix} = \sqrt{\left|\frac{(n-1)^2+n^2h^2}{(n+1)^2+n^2h^2}\right|} \times \begin{Bmatrix} \text{Amplitude of} \\ \text{incident wave} \end{Bmatrix}. \quad (25\cdot82)$$

Consequently

$$\begin{Bmatrix} \text{Intensity of} \\ \text{reflected wave} \end{Bmatrix} = \frac{(n-1)^2+n^2h^2}{(n+1)^2+n^2h^2} \times \begin{Bmatrix} \text{Intensity of} \\ \text{incident wave.} \end{Bmatrix} \quad . \ (25\cdot83)$$

§ 25·9. Reflexion at Metallic Surfaces

The quantities n and h in formula $(25\cdot83)$ can be determined in the case of waves for which τ is within the limits of the visible spectrum by optical experiments.   In this way Drude and others have found nh in the case of metals to be great by comparison with n, and formula $(25\cdot83)$ in consequence gives a high reflecting power.   This high reflecting power is the cause of the characteristic metallic lustre of metals.   If an empty test-tube (i.e. one containing air only) be immersed in water, it will be noticed that the parts of its surface seen by totally reflected light show this kind of lustre.   According to Drude, nh has the value $3\cdot67$ while n is equal to $0\cdot18$ in the case of silver and sodium light.   This makes the reflecting power of the silver more than $0\cdot95$ when we substitute in $(25\cdot83)$.   It is when we combine $(25\cdot82)$ and $(25\cdot83)$ with the formulae $(25\cdot192)$ that we meet with difficulties.   If we take the case of silver and sodium light, we find $(25\cdot192)$ that K turns out to be negative when we substitute the appropriate values of n and h.   According to Drude, a negative dielectric constant is meaningless.   While this view cannot be upheld (cf. § 26·4) the theory of § 25·1 certainly implies a positive value of K.   The second of the equations $(25\cdot192)$ also does not exhibit any relationship whatever to the observed facts in the case of silver, mercury and other metals if light in the visible spectrum is used.   In a formal way the electromagnetic theory, so far as we have developed it, accords well with the phenomena observed ; but we cannot calculate the optical constants from the electrical properties of the medium ; not at all events when we are dealing with wave-lengths within the limits of the visible spectrum.   If, however, we have to do with wave-lengths as long as those of the residual radiation (§ 26) of Rubens, we find (according to Hagen and Rubens [1]) very good accordance between the value of the reflecting power calculated from electrical data and that measured experimentally.   It is reasonably probable

[1] *Ann. d. Physik*, II, 1903, p. 873.

that $K/\sigma\tau$ is very small, and consequently h is nearly unity **(25·191)**. Hence by **(25·192)**

$$\frac{A\sigma\tau}{K_0} = n^2.$$

Substituting in **(25·83)** we get for the reflecting power

$$R = \frac{2n^2 - 2n + 1}{2n^2 + 2n + 1},$$

and as

$$n^2 \gg n \gg 1,$$

this becomes

$$R = \frac{1 - \dfrac{1}{n}}{1 + \dfrac{1}{n}} = 1 - \frac{2}{n}$$

to a sufficiently good approximation. Consequently

$$R = 1 - 2\sqrt{\left.\frac{K_0}{A\sigma\tau}\right|}, \quad . \quad . \quad . \quad . \quad \textbf{(25·9)}$$

a formula enabling the reflecting power to be calculated from the period (or wave-length in air or *in vacuo*) of the waves, and the electrical conductivity of the material. If we use ordinary electrostatic units, A and K_0 become each equal to unity and **(25·9)** consequently becomes :

$$R = 1 - 2/\sqrt{\sigma\tau}|. \quad . \quad . \quad . \quad \textbf{(25·91)}$$

The reflecting power is all the greater, the greater the conductivity and the longer period of the waves.

BIBLIOGRAPHY

In addition to the treatises of Clerk Maxwell, A. G. Webster and Clemens Schäfer, the following works may be mentioned :

CLERK MAXWELL : Scientific Papers.

O. HEAVISIDE : Electromagnetic Theory.

P. DRUDE : Physik des Aethers. (Enke, Stuttgart.) Lehrbuch der Optik. (Hirzel, Leipzig.)

J. J. THOMSON : Elements of the Mathematical Theory of Electricity and Magnetism. (Cambridge.)

F. B. PIDDUCK : A Treatise on Electricity. (Cambridge.)

S. G. STARLING : Electricity and Magnetism. (Longmans.)

CHAPTER IX

SIMPLE ELECTRON THEORY

§ 26. Origins of the Electron Theory

THE phenomena of electrolysis, in particular the laws discovered by Faraday a century ago, suggested that electrical conduction through solutions of salts, acids and similarly constituted materials (electrolytes) was effected by **ions,** charged particles consisting of single atoms, or of groups of atoms, such as H, Na, Cu, Ag, etc. (positively charged), and Cl, SO_4, NO_3 (negatively charged). The proportionality of **electrochemical equivalent** and **chemical equivalent** suggested the existence of a definite elementary charge, **atomic** in a literal sense as well as in the sense of being associated with an **atom.** Hydrogen, sodium, silver and other chemically univalent ions were thought to carry one (positive) elementary charge ; magnesium, calcium, barium to carry two such charges. Similarly, a chlorine ion and an NO_3 ion were thought to carry one negative elementary charge, an SO_4 ion two, and so on. These views have only been modified to the extent that an atom, e.g. a sodium atom, is believed to contain many of these elementary charges, its **ionic charge** being their algebraic sum. The existence of this elementary charge in air and gases when in an electrically conducting state has been demonstrated beyond any doubt, more especially by the experimental methods of C. T. R. Wilson, who discovered that in a sufficiently supersaturated atmosphere, water drops form round individual ions. This discovery also led to the most important of the experimental procedures for measuring this charge, which, according to the measurements of Millikan, amounts to $4 \cdot 774 \times 10^{-10}$ ordinary electrostatic units. The term **electron** was introduced by Johnstone Stoney (1874) as a name for this natural unit of electricity ; but the usage of later times has assigned the term ' electron ' to a special kind of ion (J. J. Thomson's **corpuscle**) first recognized in the streams of ions constituting the **cathode rays.** The most striking feature of these ions, or **electrons**

as we shall henceforth call them, is the possession of a unique ratio of charge to mass, and hence a common mass (about 1/1845 of that of a hydrogen atom [1]), whatever the material or materials may be in which they originate. We have therefore in electrons a constituent common to all atoms. This discovery led to the **electron theory of matter** which, in its extreme form, identified all mass with the inertia—to be explained in the following section—which is necessarily associated with an electric charge, and sought to build up physical theory on a purely electrical basis. This extreme form of the electron theory has long since been abandoned, not so much on account of the serious (and indeed insuperable) difficulty of constituting a charged particle without introducing *non-electrical forces*, as on account of the great possibilities and potency in the line of development which culminated in the theory of relativity.

§ 26·1. MASS ASSOCIATED WITH A SLOWLY MOVING CHARGE

We have seen **(22·61)** that the force exerted on a charged particle in a magnetic field is expressed by

$$\mathbf{F} = \frac{4\pi A}{a} e[\mathbf{v}, \mathbf{D}_m],$$

where e is the charge, $\mathbf{v}$ is the velocity of the particle and $\mathbf{D}_m$ the magnetic displacement in the field. Let the displacement, $\mathbf{D}_m$, be due to a point (north) pole of strength, m (Fig. 26·1) distant $\mathbf{r}$ from the charged particle at O. Then

$$\mathbf{D}_m = \frac{m}{4\pi \mathbf{r}^2},$$

and the force, $\mathbf{F}$, can be expressed numerically by

$$\mathbf{F} = \frac{A}{a} \frac{em\mathbf{v}}{\mathbf{r}^2} \sin \psi.$$

By Newton's law of the equality of action and reaction this equation also expresses numerically the force exerted on the point pole, m, by the moving particle. We shall write it in the form :

$$\mathbf{F} = \frac{A}{a} \frac{me\mathbf{v}}{\mathbf{r}^2} \sin \theta. \quad . \quad . \quad . \quad \mathbf{(26 \cdot 1)}$$

---

[1] We are referring here to the mass associated with low velocities.

Consequently the intensity of the magnetic field of the moving charged particle at the point m is expressed by

$$\frac{A}{a}\frac{ev}{r^2}\sin\theta, \qquad . \quad . \quad . \quad . \quad (26\cdot11)$$

and its direction is downwards from the plane of the paper (Fig. 26·1). The magnetic energy per unit volume is expressed by

$$\frac{\mu H^2}{8\pi A},$$

where H means the intensity $(26\cdot11)$, and μ is the permeability. Consequently we may write for the energy per unit volume:

$$\frac{A\mu}{8\pi a^2}\frac{e^2v^2}{r^4}\sin^2\theta. \qquad . \quad . \quad . \quad (26\cdot12)$$

Now suppose the particle to be a small sphere with the charge,

Fig. 26·1

e, uniformly distributed over its surface. Considerations similar to those which demonstrate that the electric field intensity in the neighbourhood of such a sphere at rest is, at points outside it, the same as it would be if all the charge were concentrated at its centre, and zero within the sphere, lead to identical conclusions in the present case. We therefore apply the formula $(26\cdot12)$ for exterior points, and take the energy to be zero within the sphere. The whole of the magnetic energy due to the motion of the sphere is therefore

$$\frac{A\mu e^2 v^2}{8\pi a^2}\int_R^\infty\int_0^\pi\int_0^{2\pi}\frac{dr}{r^2}\sin^3\theta\ d\theta\ d\phi,$$

where ϕ is an azimuthal angle. On integration we get for the magnetic energy:

$$\frac{A\mu e^2 v^2}{3a^2R}. \qquad . \quad . \quad . \quad . \quad (26\cdot13)$$

This energy is contributed by the *motion* of the sphere. It is in fact the **kinetic energy of the charge.**

One important point concerning the deduction of $(26\cdot13)$ should be emphasized. **It is only valid for small velocities,**

i.e. for velocities small compared with that of electromagnetic waves. We have, in fact, assumed that the magnetic field due to the motion is instantly set up at all points, however distant, in the region surrounding the charge. A more rigorous treatment of the problem will be given in **§ 28,** meanwhile we shall equate the kinetic energy **(26·13)** to $\frac{1}{2}M\mathbf{v}^2$. where M may be described as the mass associated with the charge. Therefore

$$M = \frac{2A\mu e^2}{3a^2R}. \quad . \quad . \quad . \quad . \quad . \quad (26\cdot14)$$

If we employ old-fashioned units $(A = 1)$ of the mixed type $(a = c)$, and suppose the charge to be moving in empty space, $\mu = 1$, then

$$M = \frac{2}{3}\frac{e^2}{Rc^2}, \quad . \quad . \quad . \quad . \quad (26\cdot141)$$

or, in ordinary electromagnetic units, $(A = 1, \ a = 1, \ \mu = 1)$:

$$M = \frac{2}{3}\frac{e^2}{R}. \quad . \quad . \quad . \quad (26\cdot142)$$

§ 26·2. Dispersion of Electromagnetic Waves

Our simple theory of the propagation of waves in an isotropic insulating medium provides a unique velocity, $v = a/\sqrt{\mu K}\,|$, for electromagnetic waves, whereas experiment undoubtedly demonstrates that the velocity of light-waves (which we are forced to regard as electromagnetic waves) varies with the period of vibration in the wave. The simple theory of propagation in a *conducting* medium does indeed show **(§ 25·1)** a dependence of the velocity on the period; but it too is inadequate for periods as short as those of ordinary light waves. The simple theory in both cases approximates asymptotically to the facts as the period becomes long.

Proceeding in the spirit of the electron theory of Drude and Lorentz, we shall adopt as a picture or model of an insulating medium an assemblage of charged particles situated in otherwise empty space. Its insulating quality is secured by binding these particles to fixed points. We shall, however, leave them a certain restricted freedom of movement : any displacement of an individual particle or **ion** evoking a restoring force proportional to its displacement from the fixed point. Forces depending on the velocity of the ion we shall suppose to be very small and, for most purposes, ignorable. In its uncharged state we are bound to assume the existence of at least two classes of ions,

positive and negative, the algebraic sum of the charges in any volume being—unless it be an exceeding small volume indeed, within which only one or two ions might be situated—zero. Let us further assume the structure of the medium to be so fine that an element of volume, $dx\ dy\ dz$, contains an enormous number of ions, and is therefore uncharged as a whole. We shall attack our problem most easily if we imagine in the first instance only two classes of ions, and suppose one sort, the negative ions for example, to be associated with restoring forces so big that their motion is of no consequence.

Imagine a plane harmonic wave, the period in which is τ (angular frequency $\omega = 2\pi/\tau$) travelling through the hypothetical medium in the direction of the Z axis, its vibrations being along the line of the X axis. The ions will be set in motion under the influence of the field of the wave, and their motion will cause a periodically varying polarization of the medium. We shall investigate this polarization and deduce an expression (with the aid of **19·411**) for K. Our investigation will prove K to be a function, not only of the properties of the medium, but also of the frequency ω.

So long as the assemblage of ions is not subjected to a field, we may think of them as paired to form doublets, or dipoles, of negligible electric moment. The effect of an electric field in the X direction is to drag each positive ion some distance, x, in that direction. If N be the number of positive ions per unit volume, it is easily seen that the charge displaced through the unit area normal to the X axis is Nex. Therefore the polarization, P, is

$$P = Nex. \quad . \quad . \quad . \quad . \quad \textbf{(26·2)}$$

The product, ex, is the electric moment of a single dipole, and therefore the polarization is equal to the electric moment per unit volume.

§ 26·3.   Internal Field in a Dielectric Medium

Superposed on the electric field of the wave (which we may call the impressed field) will be one evoked by the displacement of the ions. To find out how this **internal field**, at any point, depends on the polarization, imagine a small sphere (large enough of course to contain an enormous number of dipoles) described round the point as centre. We may regard the internal field at its centre as due to :

(a) the dipoles within the spherical volume and
(b) polarization charges which would appear over the surface of the sphere if the dipoles within it were removed.

We shall begin with (a). Let us place the origin of a system of rectangular co-ordinates at the centre of a dipole, and the X axis in the direction of its moment. We have then for the potential at a point P (Fig. 26·3 (a)),

$$V = \frac{AMx}{K_0 r^3}, \quad . \quad . \quad . \quad . \quad (26·3)$$

where M is the moment of the dipole, K_0 is the dielectric constant of empty space and r is the distance of the point, P, from the
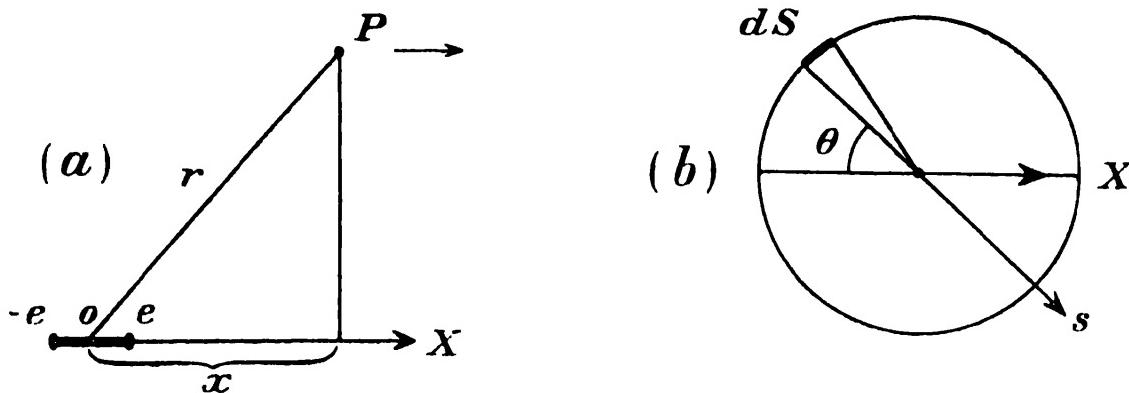

FIG. 26·3

centre of the dipole, the latter being assumed to be very short by comparison with r (20·71). The component in the X direction of the electric field intensity at the point, P, due to the dipole is

$$-\frac{\partial V}{\partial x} = \frac{AM}{K_0 r^3}\left(\frac{3x^2}{r^2} - 1\right). \quad . \quad . \quad . \quad (26·31)$$

If now we shift the origin from the dipole to P, without changing the directions of the axes, this formula will still remain valid, because x appears in it squared. Now suppose this point P to be the centre of the spherical region. Under the influence of the field we are assuming, the axes of all the dipoles are parallel to one another, and their moments have all the same direction, namely that of the X axis. The X component of the field intensity (due to them) at the centre of the sphere may therefore be written :

$$f_x = \frac{AM}{K_0}\sum \frac{1}{r^3}\left(\frac{3x^2}{r^2} - 1\right), \quad . \quad . \quad . \quad (26·32)$$

the summation being extended over all the dipoles in the spherical region. Since the symbols x and r refer to the positions of the centres of the dipoles only, it is obvious that f_x may equally well be written :

$$f_x = \frac{AM}{K_0}\sum \frac{1}{r^3}\left(\frac{3y^2}{r^2} - 1\right). \quad . \quad . \quad (26·321)$$

Similarly, we may write

$$f_x = \frac{AM}{K_0}\sum\frac{1}{r^3}\left(\frac{3z^2}{r^2} - 1\right). \quad . \quad . \quad . \quad \text{(26·322)}$$

When we add together the expressions (26·32), (26·321) and (26·322), we find that $3f_x$, and consequently f_x, is identically zero, and we may similarly (or by the argument from symmetry) show that the components f_y and f_z are likewise zero.

So we are left with the contribution (b) due to the unbalanced polarization charges left over the spherical surface when the dipoles in the interior, which as we have seen contribute nothing, are removed. The charge on the element dS (Fig. 26·3 (b)) is equal to $PdS\cos\theta$, the polarization having the direction of the X axis. This will give rise to a field intensity at the centre of the sphere equal to

$$\frac{APdS\cos\theta}{K_0R^2}, \quad . \quad . \quad . \quad . \quad \text{(26·33)}$$

in the direction s. The symbol A represents our familiar units constant, and K_0 the dielectric constant of empty space. The component of the intensity in the X direction is therefore

$$\frac{APdS\cos^2\theta}{K_0R^2}. \quad . \quad . \quad . \quad \text{(26·34)}$$

Substituting for dS the usual expression, $R^2\sin\theta\,d\theta\,d\phi$, and integrating over the ranges 0 to π for θ, and 0 to 2π for ϕ, we at once obtain

$$\frac{4\pi AP}{3K_0}. \quad . \quad . \quad . \quad . \quad \text{(26·35)}$$

The remaining components are easily shown, either by the foregoing method, or by reasoning from symmetry, to be zero. The expression (26·35) represents what we shall call the **internal field.** It is the field intensity evoked by the polarization. The total field intensity at any point in the medium will therefore be :

$$\mathcal{E} + \frac{4\pi AP}{3K_0},$$

where $\mathcal{E}$ represents the impressed field, and since by (19·411)

$$P = \frac{(K - K_0)}{4\pi A}\mathcal{E},$$

the total field intensity may be written :

$$\frac{(K + 2K_0)}{3K_0}\mathcal{E}. \quad . \quad . \quad . \quad \text{(26·36)}$$

It is instructive to compare this formula with (20·98).

§ 26·4. Dispersion in an Insulating Medium

The equation of motion of an individual ion, on the hypotheses set out above, is :

$$m\frac{d^2x}{dt^2} = -\eta x + \frac{K + 2K_0}{3K_0}e\mathcal{E}. \quad . \quad . \quad . \quad (26\cdot4)$$

In this equation η represents the restoring force per unit displacement, and $\mathcal{E}$ is the electric field intensity impressed on the medium by the plane wave. We have ignored the contribution of the magnetic field in the wave to the force on the ion. This is justifiable, as a study of the expression (22·63) shows. In this formula $\mathcal{E}$ and H are measured in mixed units, and in these units $\mathcal{E}$ and H in an electromagnetic wave in empty space are numerically equal. The term involving H is, however, multiplied by the very small factor v/c. In the case before us the circumstances are slightly different ; but it is obvious that the contribution of the magnetic field to the force on the ion is bound to be very small by comparison with that due to the electric field—unless indeed we have enormous field intensities, so that v/c ceases to be negligible by comparison with unity. In equation (26·4) we have also ignored dissipative forces.

If we divide through by m, and write ω_0^2 for η/m, and if we further write for $\mathcal{E}$ the expression

$$Ee^{i\omega(t-z/v)}, \quad . \quad . \quad . \quad . \quad (26\cdot401)$$

where E is a positive real constant,[1] and v is the velocity of the wave, we get

$$\frac{d^2x}{dt^2} + \omega_0^2 x = \frac{(K + 2K_0)}{3K_0 m}eEe^{i\omega(t-z/v)}.$$

We may simplify this if we bear in mind that we are investigating the motion of a particular ion and that consequently z is a constant. Therefore

$$Ee^{i\omega(t-z/v)} = \text{constant} \times Ee^{i\omega t},$$

and we may suppose the constant to be contained in E. Thus we have :

$$\frac{d^2x}{dt^2} + \omega_0^2 x = \frac{(K + 2K_0)}{3K_0 m}eEe^{i\omega t}. \quad . \quad . \quad (26\cdot41)$$

---

[1] The expression for the harmonic wave has properly the form :

$$E \cos \omega(t - z/v),$$

or $$E \sin \omega(t - z/v),$$

with a suitable choice of the arbitrary constant in the phase. For a justification of the use of the exponential see § 23·7.

The method we have already adopted (§ 23·7) for an equation of this form yields the particular solution :

$$x = \frac{(K + 2K_0)e}{3K_0 m(\omega_0^2 - \omega^2)} E e^{i\omega t},$$

or

$$x = \frac{(K + 2K_0)e}{3K_0 m(\omega_0^2 - \omega^2)} \mathcal{E}. \quad . \quad . \quad . \quad (26\cdot42)$$

We shall adopt this expression for x, notwithstanding the fact that it only represents a particular solution, on the grounds that even very small dissipative terms in the differential equation would cause the rest of the solution to be multiplied (as in 23·73) by an exponential factor which would practically wipe it out completely after the lapse of a finite interval of time, while such very small dissipative terms would not *appreciably* modify [1] the particular expression which endures.

If now we multiply both sides of (26·42) by Ne we get on the left the polarization, P (26·2), and we shall naturally replace it by

$$\frac{(K - K_0)}{4\pi A} \mathcal{E},$$

in accordance with (19·411). Thus we have

$$\frac{(K - K_0)}{4\pi A} \mathcal{E} = \frac{(K + 2K_0)Ne^2}{3K_0 m(\omega_0^2 - \omega^2)} \mathcal{E},$$

or

$$\frac{K - K_0}{K + 2K_0} = \frac{4\pi A Ne^2}{3K_0 m(\omega_0^2 - \omega^2)}. \quad . \quad . \quad (26\cdot43)$$

If instead of assuming only one class of movable ions, we assume the existence of two or more classes, and distinguish them and their associated constants by numeral subscripts, we shall find (26·43) amplified in the following way :

$$\frac{K - K_0}{K + 2K_0} = \frac{4\pi A}{3K_0} \sum_s \frac{N_s e_s^2}{m_s(\omega_s^2 - \omega^2)}, \quad . \quad . \quad (26\cdot44)$$

where N_s is the number of ions of the s class per unit volume, e_s is the charge on such an ion, m_s is its mass and ω_s its natural (angular) frequency. The symbol ω (without subscript) represents, as before, the angular frequency of vibration in the wave.

---

[1] Except in the neighbourhood of resonance.

Since

$$\omega^2 = 4\pi^2/\tau^2 \text{ and } c^2\tau^2 = \lambda^2,$$

where τ means period and λ the corresponding wave-length *in vacuo*, we have

$$\dot\omega^2 = \frac{4\pi^2 c^2}{\lambda^2},$$

$$\omega_s^2 = \frac{4\pi^2 c^2}{\lambda_s^2}.$$

Here λ is the wave-length (*in vacuo*) of the wave we are supposing to traverse the insulating medium, λ_s is the wave-length (*in vacuo*) of waves which have a frequency equal to the natural one ω_s. We may therefore write instead of (26·44),

$$\frac{K - K_0}{K + 2K_0} = \frac{A\lambda^2}{3\pi K_0 c^2} \sum_s \frac{N_s e_s^2 \lambda_s^2}{m_s(\lambda^2 - \lambda_s^2)}. \qquad . \qquad . \quad (26·45)$$

It is quite clear that the permeability of our hypothetical insulator is identical, or nearly so, with that of empty space. Hence its refractive index, n, is expressed by

$$n = \sqrt{\frac{K}{K_0}}.$$

Substituting in (26·45) we get :

$$\frac{n^2 - 1}{n^2 + 2} = \frac{A\lambda^2}{3\pi K_0 c^2} \sum_s \frac{N_s e_s^2 \lambda_s^2}{m_s(\lambda^2 - \lambda_s^2)}, \qquad . \qquad . \quad (26·46)$$

a formula which expresses the way in which the refractive index of the medium depends on the wave-length (as measured *in vacuo*) of the waves.

§ 26·5. DEPENDENCE OF REFRACTIVE INDEX ON STATE OF AGGREGATION

If we make the hypothesis that the charge e_s on the s type of ion does not vary with the state of aggregation of the medium, and that the restoring force acting on such an ion is a property of the molecule and therefore also (when no dissociation occurs on change of state) independent of state of aggregation, we are led to the conclusion that for waves of given λ (say that of one of the sodium D lines) only the factor N_s in (26·46) will change when the medium changes from solid to liquid, or from liquid

to vapour. And N_s will obviously be proportional to its density, ρ. It follows therefore that

$$\frac{n^2 - 1}{n^2 + 2} = C\rho, \quad . \quad . \quad . \quad . \quad . \quad \textbf{(26·5)}$$

where C is a constant characteristic of the substance. This relationship between the refractive index and the density of the substance in its different states of aggregation was discovered independently by Lorenz and H. A. Lorentz (1880). It is in extraordinarily close agreement with observations in the case of many substances.

§ 26·6. Examination of the Dispersion Formula

The formula **(26·46)** may be written :

$$\frac{n^2 - 1}{n^2 + 2} = \alpha_1\frac{\lambda^2}{\lambda^2 - \lambda_1{}^2} + \alpha_2\frac{\lambda^2}{\lambda^2 - \lambda_2{}^2} + \ldots \quad . \quad \textbf{(26·6)}$$

in which the wave-lengths λ_1, λ_2, . . . corresponding to the natural periods of the different groups of ions may be taken to be in the order of increasing wave-length, i.e. $\lambda_1 < \lambda_2 < \lambda_3$ . . . and so on, and in which α_1, α_2. . . . are positive constants. If we denote

$$\alpha_s \frac{\lambda^2}{\lambda^2 - \lambda_s{}^2} \quad \text{by} \quad L_s,$$

we may write instead of **(26·6)**

$$\frac{n^2 - 1}{n^2 + 2} = L_1 + L_2 + \ldots + L_s + \ldots = L, \quad \textbf{(26·61)}$$

or

$$n^2 = \frac{1 + 2L}{1 - L}. \quad . \quad . \quad . \quad . \quad \textbf{(26·62)}$$

We see at once, on inspecting **(26·6)** that as λ approaches zero n^2 approaches unity. The theory thus accounts for the inappreciable refraction of extremely short waves (X-rays and γ-rays). As λ increases from zero, n^2 falls below unity, getting smaller and smaller, and eventually becoming *negative*. This is obvious so long as $\lambda < \lambda_1 < \lambda_2$ . . . because then the right-hand side of **(26·46)** is negative. As λ approaches the shortest natural wave-length λ_1, L approaches ∞ on the negative side, and **(26·62)** n^2 approaches -2. As λ increases beyond λ_1, while still in its immediate neighbourhood,

$$n^2 = \frac{1 + 2L_1}{1 - L_1}$$

approximately, and L_1 is a large, but decreasing, positive number. As L_1 decreases till it approaches $+1$ from above, n^2 approaches $-\infty$, and becomes $+\infty$ as L_1 drops below $+1$. With still further increase of λ the curve will descend as indicated by *efg* in Fig. 26·6. This part represents what was formerly regarded as the normal type of dispersion—such as that of visible light in ordinary glass. As λ now approaches the next natural wavelength, λ_2, the curve *ghij* repeats approximately the part *bcde*, cutting the vertical line, λ_2, at a distance 2 below zero. If there
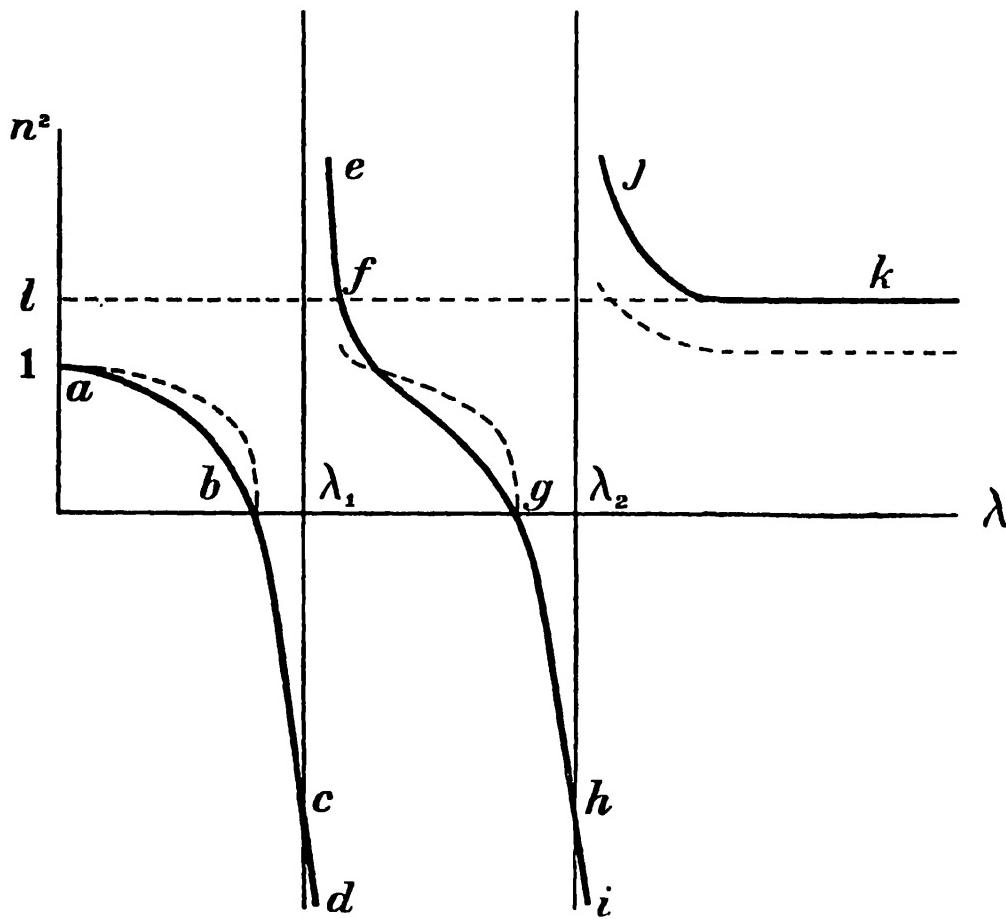

Fig. 26·6

are only *two* sets of movable ions, it will eventually approach asymptotically to horizontality somewhere beyond *k*. The special value of n^2 (represented by the point *l*) corresponding to this, being the electrostatic value of K, expressed in E.S. units.

So long as n^2 is positive n obviously means the observed refractive index. But the negative values of n^2 require interpretation. The expression **(26·401)** for the wave—since $v = c/n$ —may be written :

$$\mathcal{E} = E e^{i\omega\left(t - \frac{nx}{c}\right)}.$$

When n^2 is negative we shall write for n

$$n = i\beta,$$

where β is *real*. Thus

$$\mathcal{E} = Ee^{\frac{\omega\beta z}{c}}e^{i\omega t},$$

or, taking the real parts,

$$\mathcal{E} = Ee^{\frac{\omega\beta z}{c}}\cos\omega t. \quad . \quad . \quad . \quad . \quad \textbf{(26·63)}$$

This formula, it will be seen, does not represent a wave at all,[1] but a state of affairs in which every particle is in simple harmonic vibration, and all of them at a given instant in the same phase. The amplitudes, $Ee^{\omega\beta z/c}$, of these vibrations vary exponentially from one value of z to another. When the propagation, as we are supposing, is in the direction of increasing values of z, the constant β is bound to be negative. A very simple application of the energy principle will make this evident. And since n^2 approaches $-\infty$ at a certain point (Fig. 26·6) in the neighbourhood of the natural wave-length, λ_s, it follows that here too β approaches $-\infty$. Such plane waves, therefore, falling perpendicularly on a plane face of the medium, will set up a state vibration in the surface ; but will not travel onwards into the medium. They will in fact suffer complete reflexion. This explains the **selective reflexion** which is found to be associated with anomalous dispersion and also the **residual radiation (Restsrahlung)** observed by Rubens and his associates. There is then a narrow region in the neighbourhood of each λ_s where n^2 is negative and where in consequence the material strongly reflects. Outside such regions n^2 is positive and progressive waves can travel through the medium. This state of affairs can exist quite close to the natural wave-lengths λ_s, and since the disturbance impressed by the wave has here almost the same frequency as the natural one of the class of ions concerned, we have a state very closely approaching resonance, and consequent strong absorption, a familiar phenomenon associated with anomalous dispersion.

The number n in the foregoing theory of dispersion is either real and positive or purely imaginary. In the former case n represents the observable refractive index. Its values for different wave-lengths are roughly indicated by the course of the dotted curves in Fig. 26·6. In the neighbourhood of one of the natural wave-lengths, λ_s, n springs discontinuously from

---

[1] We might describe it as a wave with an infinite phase velocity.

0 to $+\infty$ if we take (26·6) to be exact. In the immediate neighbourhood of resonance, however, i.e. in the neighbourhood where $(\lambda - \lambda_s)$ or $(\omega - \omega_s)$ is very small or zero, the dissipative term, to which reference has been made, cannot be ignored. If we introduce the term $2\alpha\, dx/dt$ in the left-hand side of (26·41), α being a small positive constant, equation (26·43) will in consequence become

$$\frac{K - K_0}{K + 2K_0} = \frac{4\pi A N e^2}{3 K_0 m (\omega_0^2 - \omega^2 + 2i\alpha\omega)},$$

so that in reality our theory allows the observable refractive index to range between very small and very great values, but does not permit the physically impossible discontinuity just mentioned.

For information about the application of the classical electromagnetic theory to dispersion in conducting or absorbing media the reader is referred to Drude's *Lehrbuch der Optik* and other works mentioned in the bibliography at the end of Chapter XII.

§ 26·7. The Zeeman Phenomenon

When a source of light, such as a bunsen flame containing a metallic vapour (a sodium flame for example), or a tube containing a rarefied gas through which an electric current is passed, is placed in the strong magnetic field between the poles of an electromagnet and one of the spectral lines observed by some spectroscopic device of sufficiently high resolving power, it is found that the original line splits into a number of lines. In the simplest case there are three of them, or only two, according to the direction of the line of observation relatively to that of the magnetic field. The light associated with the component lines is polarized in a characteristic way described below. The phenomenon seems to have been first observed by P. Zeeman.[1] We shall describe here the theory of the effect as given by H. A. Lorentz, on account of its historical interest and its close relationship to the theory of dispersion described in §§ 26·2–26·6.

We suppose the monochromatic radiation emitted by some excited material—for example, the light of the yellow line of helium gas excited by an electric discharge—to be due to the simple harmonic vibrations of an ion. Again we neglect dissi-

[1] P. Zeeman, *Phil. Mag.*, 43, p. 226, and 44, p. 255, 1897.

pative forces and write the equations of motion of the ion in the form :

$$m\frac{d^2x}{dt^2} = -\eta x + \frac{4\pi A c}{a}[\mathbf{v}, \mathbf{D}_m]_x,$$
$$m\frac{d^2y}{dt^2} = -\eta y + \frac{4\pi A e}{a}[\mathbf{v}, \mathbf{D}_m]_y, \quad . \quad . \quad (26\cdot7)$$
$$m\frac{d^2z}{dt^2} = -\eta z + \frac{4\pi A e}{a}[\mathbf{v}, \mathbf{D}_m]_z.$$

The terms containing the vector product $[\mathbf{v}, \mathbf{D}_m]$ represent the contribution of the magnetic field to the force on the ion. If we employ the mixed type of units, the permeability, which we take to be that of empty space, becomes unity, while $a = c$. We shall further suppose the magnetic field, $\mathbf{H}$, to have the direction of the Z axis. Therefore

$$[\mathbf{v}, \mathbf{D}_m] = \frac{1}{4\pi A}[\mathbf{v}, \mathbf{H}].$$

Remembering this, and dividing equations (26·7) by m and writing, as before, $\omega_0{}^2$ for η/m, we get :

$$\frac{d^2x}{dt^2} + \omega_0{}^2 x = \frac{e}{mc}H\frac{dy}{dt},$$
$$\frac{d^2y}{dt^2} + \omega_0{}^2 y = -\frac{e}{mc}H\frac{dx}{dt}, \quad . \quad . \quad . \quad (26\cdot71)$$
$$\frac{d^2z}{dt^2} + \omega_0{}^2 z = 0,$$

in which H is the magnetic field intensity, or its Z component, since we are supposing it to be directed along the Z axis. The third equation is unmodified by the presence of the field and consequently the motions of the ion along the line of the field are just those it would execute in the absence of a field. Whatever radiations the ion may emit therefore in consequence of the presence of the field, it will continue to emit radiation of the original unmodified frequency ω_0. Only we notice that this will be polarized in the sense corresponding to motions along the line of the field.

The first and second equations (26·71) cannot be treated separately from one another. We get a particular solution of them by writing

$$x = Me^{i\omega t},$$
$$y = Ne^{i\omega t}, \quad . \quad . \quad . \quad . \quad (26\cdot72)$$

and substituting in (26·71). We thus find that the constants M and N are connected by the relations :

$$\left.\begin{array}{c} (\omega_0{}^2 - \omega^2)M = i\omega\dfrac{eH}{mc}N, \\[3mm] (\omega_0{}^2 - \omega^2)N = - i\omega\dfrac{eH}{mc}M. \end{array}\right\} \quad \text{. . . . } \textbf{(26·73)}$$

Hence

$$\frac{M}{N} = - \frac{N}{M},$$

or

$$M = \pm iN. \quad \text{. . . . . } \textbf{(26·74)}$$

The substitution of $+ iN$ for M in either of the equations (26·73) gives for ω the quadratic equation :

$$\omega^2 + \frac{eH}{mc}\omega - \omega_0{}^2 = 0 ; \quad \text{. . . . } \textbf{(26·75)}$$

and if we replace M by $- iN$ we get :

$$\omega^2 - \frac{eH}{mc}\omega - \omega_0{}^2 = 0. \quad \text{. . . . } \textbf{(26·751)}$$

The solution of the former of these is

$$\omega = - \frac{eH}{2mc} \pm \omega_0\sqrt{1 + \frac{e^2H^2}{4m^2c^2}\cdot\frac{1}{w_0{}^2}}.$$

Experiment indicates that the frequencies, ω, differ extremely little from ω_0, and therefore $eH/2mc.\omega_0$ is a very small number, and $e^2H^2/4m^2c^2\omega_0{}^2$ entirely negligible. The physically significant root of (26·75) is therefore

$$\omega' = \omega_0 - \frac{eH}{2mc}. \quad \text{. . . . } \textbf{(26·752)}$$

Similarly, the physically significant root of (26·751) is

$$\omega'' = \omega_0 + \frac{eH}{2mc}. \quad \text{. . . . } \textbf{(26·753)}$$

In addition to the original frequency ω_0, we have therefore the two frequencies, ω' and ω'', which we may conveniently write in the form :

$$\left.\begin{array}{c} \omega' - \omega_0 = - \dfrac{eH}{2mc}, \\[3mm] \omega'' - \omega_0 = + \dfrac{eH}{2mc} ; \end{array}\right\} \quad \text{. . . } \textbf{(26·76)}$$

or in the form :

$$\left.\begin{array}{l} \nu' - \nu_0 = -\dfrac{eH}{4\pi mc}, \\[2mm] \nu'' - \nu_0 = +\dfrac{eH}{4\pi mc}, \end{array}\right\} \quad . \quad . \quad . \quad \textbf{(26\cdot761)}$$

if ν', ν'', ν_0 represent frequencies expressed in *vibrations* (*or revolutions*) *per unit time*, since

$$\omega = 2\pi/\tau = 2\pi\nu.$$

With the frequency ω' is associated the amplitude $M = + iN$. Substituting in **(26·72)** and replacing the letter N by A, we have :

$$x = iAe^{i\omega't},$$
$$y = Ae^{i\omega't},$$

or

$$x = Ae^{i(\omega't+\pi/2)},$$
$$y = Ae^{i\omega't},$$

since

$$i = e^{i\pi/2}.$$

The arbitrary constant, A, may be real or complex, and consequently may be written :

$$A = A_0 e^{i\delta},$$

where A_0 is real and positive and δ is real. Therefore the solution becomes :

$$x = A_0 e^{i(\omega't+\pi/2+\delta)},$$
$$y = A_0 e^{i(\omega't+\delta)}.$$

This solution includes the real one :

$$\left.\begin{array}{l} x = A_0 \cos(\omega't + \pi/2 + \delta), \\ y = A_0 \cos(\omega't + \delta). \end{array}\right\} \quad . \quad . \quad . \quad \textbf{(26\cdot77)}$$

Similarly, we obtain for the second solution :

$$\left.\begin{array}{l} x = B_0 \cos(\omega''t - \pi/2 + \varepsilon), \\ y = B_0 \cos(\omega''t + \varepsilon), \end{array}\right\} \quad . \quad . \quad \textbf{(26\cdot771)}$$

where B_0 is any real and positive constant and ε any real number.

It is easy to see that the former of the two solutions represents a motion of the ion in a circle of radius A_0, and in the sense in which the current is flowing round the coil of the magnet producing the field H. The associated radiation is therefore circularly polarized in this sense, if viewed along the line of the magnetic field ; while the radiation associated with the latter solution is circularly polarized in the opposite sense. If the

radiation is viewed along a line through the source and per-
pendicular to the field, the circular motion projects into a linear
one, and the radiation is plane polarized, the vibrations being
perpendicular to the field or to the Z axis. Instead of the
original single spectral line, therefore, three lines are seen : one
of them in the position of the original line and associated with
plane polarized light, if viewed perpendicularly to the field, the
corresponding **plane of polarization** being defined as any plane
perpendicular to the field. The other two lines are on either side
of the original position and, by **(26·761)**, equally distant from
it as measured in terms of frequency. If viewed perpendicularly
to the field, they are also associated with plane polarized light,
the plane of polarization being one containing the line of the field
and the line connecting the source and the observer. If the
observer looks along the line of the field only the two outer lines
are seen, the associated polarization being circular as already
explained. The central line is absent because the corresponding
motions of the ion have no component perpendicular to the
field. All this is confirmed by observation, at all events when
the field is strong enough.

The conventional definition of the plane of polarization
identifies it with the plane of incidence in the case of light
polarized by reflexion at a glass plate (**§ 25·7**), and the investiga-
tion in **§ 25·6** indicates that **the electric vibrations in electro-
magnetic waves, or in light waves, are perpendicular to
the conventional plane of polarization.** This is confirmed by
observations of the Zeeman phenomenon, which establish the
identity of the plane of polarization as defined above with the
conventional one (cf. **§ 25·7**).

§ 26·8. NATURE OF THE EMITTING ION

How can we decide which of the two extreme frequencies
should be identified with ω', and which with ω'' ? We have
seen that each of these is associated with a definite sense of
rotation of the ion in its circular path, and we can therefore
decide which of the frequencies is ω' by noting the sense of the
circular polarization of the radiation. Observation shows that
$\omega' - \omega_0$ in **(26·76)** is **positive**. This means that e, the charge,
on the ion is **negative**. Measurement of the difference $\omega' - \omega_0$,
or of $\omega'' - \omega_0$, and of the field intensity, H, enable the ratio
e/m to be evaluated. Zeeman found from observations of the
sodium D lines the value

$$e/m = -1 \cdot 6 \times 10^7$$

ordinary E.M. units per gram. This is near the value obtained from measurements carried out on cathode rays and β particles, and we conclude, as Zeeman and Lorentz did, that the radiating ion is an electron. The ratio of charge to mass for a slowly moving electron turns out to be $-1 \cdot 77 \times 10^7$ ordinary E.M. units per gram according to later and more-refined observations, not only of the Zeeman effect, but also of the behaviour of electrons moving through a magnetic field (cf. § 27·3).

The same results as regards frequencies and polarization are obtained if we suppose the emitting ion to be one which travels in an approximately elliptic orbit round a central nucleus attracting it in accordance with the inverse square law. The effect of a magnetic field on the motion of such an ion may be described in the following way : the ion continues to travel in an ellipse the dimensions of which are inappreciably affected by the field. The angle between a normal to the plane of the ellipse and the direction of the magnetic field remains practically constant ; while the ellipse precesses round the axial line passing through the nucleus parallel to the field, with the angular velocity $eH/2mc$. This is called the **Larmor precession.** The theory of the Larmor precession is too long to be included here, and in any case this, or any other, classical theory is totally inadequate to render an account of the more complicated types of Zeeman resolution.

CHAPTER X

ELECTRON THEORY

§ 26·9. Stress Momentum and Energy in Electro-magnetic Fields

IT has been explained in **§ 20** that the forces exerted on charged bodies in an electrostatic field can be described in terms of a stress tensor This kind of description can be extended to include electromagnetic fields in general. Before proceeding with this extension, it will be helpful to review the main features of the investigation of the stress tensor of **§ 20.** We notice, firstly, that the forces in question are those exerted on charges or charged bodies in the field. The resultant force on any body is therefore zero when it is uncharged. The stresses do not of course necessarily vanish in such a case, and may exist in regions where charges are absent. Indeed, the stresses in an uncharged dielectric interest us no less than do stresses elsewhere. Secondly, the stresses differ from those met with in the study of elasticity. They are stresses in a sense which is defined by the way in which they enter the formula expressing the force, and are not directly associated with any deformation of the medium at all (cf. the illustration at the end of **§ 20·1).** The more general expressions which we are going to derive, including as they do cases which are not static, may be anticipated to contain terms which express the time rate of change of something. This something we must term a **momentum,** on the ground that its rate of change is a contribution to a force, and not on the ground that it is the product of a mass and a velocity. We shall indeed meet with a special case (**§ 27·1**) where we can associate a velocity, v, with this momentum, and we shall then call the quotient of the momentum by the velocity, **mass.**

Let us imagine a small body, with a charge density, ρ, situated in the dielectric and capable of moving through it. It will be helpful to think of the dielectric as a liquid. Instead of formula **(20)** we must obviously use **(22·62),** modified to represent the force per unit volume by the replacement of e by ρ (charge per unit volume) ; thus

$$\mathbf{f} = \rho \left\{ \boldsymbol{\varepsilon} + \frac{4\pi A}{a} [\mathbf{v}, \mathbf{D}_m] \right\}, \quad . \quad . \quad . \quad (26\cdot9)$$

where

$$\mathbf{D}_m = \frac{\mu}{4\pi A} \mathbf{H}.$$

The permeability μ we shall suppose to be a constant character-istic of the dielectric medium. We may also write

$$\mathbf{f} = \rho\left\{\mathcal{E} + \frac{1}{a}[\mathbf{v}, \mu\mathbf{H}]\right\}, \quad . \quad . \quad . \quad \textbf{(26·91)}$$

and consequently

$$\mathbf{f}_x = \rho\mathcal{E}_x + \frac{\rho v_y}{a}\mu H_z - \frac{\rho v_z}{a}\mu H_y. \quad . \quad . \quad \textbf{(26·92)}$$

The appropriate form of the electromagnetic field equations is as follows :

$$\left.\begin{aligned}
\frac{K}{a}\frac{\partial\mathcal{E}}{\partial t} + \frac{4\pi A}{a}\rho\mathbf{v} &= \operatorname{curl}\mathbf{H}, \\
-\frac{\mu}{a}\frac{\partial\mathbf{H}}{\partial t} &= \operatorname{curl}\mathcal{E}, \\
\operatorname{div}\mathcal{E} &= \frac{4\pi A}{K}\rho \\
\operatorname{div}\mathbf{H} &= 0.
\end{aligned}\right\} \quad . \quad . \quad \textbf{(26·93)}$$

By the aid of the first and third of these equations we eliminate ρ, v_y and v_z from **(26·92)** and thus obtain

$$4\pi A f_x = K\mathcal{E}_x\left(\frac{\partial\mathcal{E}_x}{\partial x} + \frac{\partial\mathcal{E}_y}{\partial y} + \frac{\partial\mathcal{E}_z}{\partial z}\right) + \mu H_z\left(\frac{\partial H_x}{\partial z} - \frac{\partial H_z}{\partial x} - \frac{K}{a}\frac{\partial\mathcal{E}_y}{\partial t}\right)$$

$$- \mu H_y\left(\frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} - \frac{K}{a}\frac{\partial\mathcal{E}_z}{\partial t}\right).$$

Therefore

$$4\pi A f_x = \frac{\partial}{\partial x}\left\{\frac{K}{2}\mathcal{E}_x{}^2 - \frac{\mu}{2}H_z{}^2 - \frac{\mu}{2}H_y{}^2\right\}$$

$$+ \frac{\partial}{\partial y}\left\{K\mathcal{E}_x\mathcal{E}_z + \mu H_x H_y\right\}$$

$$+ \frac{\partial}{\partial z}\left\{K\mathcal{E}_x\mathcal{E}_y + \mu H_x H_z\right\}$$

$$- \frac{\mu K}{a^2}\frac{\partial}{\partial t}\{a(\mathcal{E}_y H_z - \mathcal{E}_z H_y)\}$$

$$- K\mathcal{E}_y\frac{\partial\mathcal{E}_x}{\partial y} - K\mathcal{E}_z\frac{\partial\mathcal{E}_x}{\partial z} - \mu H_x\frac{\partial H_z}{\partial z}$$

$$- \mu H_x\frac{\partial H_y}{\partial y} + \frac{\mu K}{a}\mathcal{E}_y\frac{\partial H_z}{\partial t} - \frac{\mu K}{a}\mathcal{E}_z\frac{\partial H_y}{\partial t}.$$

The second and fourth of the field equations **(26·93)** enable us to express the last six terms in the form :

$$\frac{\partial}{\partial x}\left\{ -\frac{K}{2}\mathcal{E}_y{}^2 - \frac{K}{2}\mathcal{E}_z{}^2 + \frac{\mu}{2}H_x{}^2 \right\},$$

and therefore

$$f_x = \frac{\partial t_{xx}}{\partial x} + \frac{\partial t_{xy}}{\partial y} + \frac{\partial t_{xz}}{\partial z} - \frac{\partial M_x}{\partial t}, \quad . \quad . \quad . \quad \textbf{(26·94)}$$

where

$$\left.\begin{aligned}
t_{xx} &= \frac{1}{4\pi A}\{K(\mathcal{E}_x{}^2 - \tfrac{1}{2}\mathcal{E}^2) + \mu(H_x{}^2 - \tfrac{1}{2}\mathbf{H}^2)\}, \\[4pt]
t_{xy} &= \frac{1}{4\pi A}\{K\mathcal{E}_x\mathcal{E}_y + \mu H_x H_y\}, \\[4pt]
t_{xz} &= \frac{1}{4\pi A}\{K\mathcal{E}_x\mathcal{E}_z + \mu H_x H_z\}, \\[4pt]
M_x &= \frac{\mu K}{a^2}\left\{\frac{a}{4\pi A}(\mathcal{E}_y H_z - \mathcal{E}_z H_y)\right\}.
\end{aligned}\right\} \qquad \textbf{(26·95)}$$

The formula **(26·94)** for the X component of the force per unit volume, and the corresponding formulae for the remaining components, represent the force as due partly to a stress tensor, t_{xx}, t_{xy}, t_{yz}, etc., and partly to the rate of decrease of a momentum, **M**. We may therefore regard the electromagnetic field as endowed with a momentum equal to **M** per unit volume, where

$$\mathbf{M} = \frac{\mu K}{a^2} \times \frac{a}{4\pi A}[\mathcal{E}, \mathbf{H}],$$

or

$$\mathbf{M} = \mathbf{p}/u^2, \quad . \quad . \quad . \quad . \quad . \quad \textbf{(26·96)}$$

p being Poynting's vector **(25·33)** and u the (phase) velocity of electromagnetic waves in the medium. In the case of a plane and plane polarized wave in an isotropic medium

$$\mathbf{p} = \frac{a}{4\pi A}\,\mathcal{E}\mathbf{H},$$

since $\mathcal{E}$ and **H** are perpendicular to one another ; further,

$$\mathbf{H} = \sqrt{\frac{K}{\mu}}\,\bigg|\,\mathcal{E},$$

and therefore

$$\mathbf{p} = \frac{a}{\sqrt{\mu K}}\bigg|\,\frac{K\mathcal{E}^2}{4\pi A}$$

(§ 25), consequently

$$\mathbf{p} = uU,{}^{1} \quad . \quad . \quad . \quad . \quad . \quad . \quad (26\text{·}97)$$

where u is the velocity of the wave—the **phase velocity**—and U is the energy per unit volume. We conclude therefore that the electromagnetic momentum, $\mathbf{M}$, per unit unit volume is :

$$\mathbf{M} = U/u,$$

or

$$\mathbf{M}u = U. \quad . \quad . \quad . \quad . \quad . \quad (26\text{·}98)$$

A case of special interest, from the point of view of the electron theory more particularly, is that in which the dielectric is empty space, or the free aether. We then have $u = c$, and consequently

$$\left.\begin{array}{l} \mathbf{p} = cU, \\ \mathbf{M}c = U. \end{array}\right\} \quad . \quad . \quad . \quad . \quad . \quad (26\text{·}99)$$

§ 27. Pressure of Electromagnetic Waves

The components, t_{xx}, t_{xy}, t_{yz}, etc., of the stress tensor **(26·95)** differ only from those of **§ 20** by the addition of magnetic terms analogous to the electrical ones. In the case of the field of a plane wave travelling in the X direction, $\mathcal{E}_x$ and H_x are zero and t_{xx} reduces to

$$t_{xx} = -\frac{1}{8\pi A}\{K\mathcal{E}^2 + \mu\mathbf{H}^2\},$$

or $$t_{xx} = -U.$$

If the wave falls perpendicularly on a plane surface which is either perfectly absorbing, or which reflects a plane wave backwards along the X direction, a tension equal to $-U$, that is to say a **pressure** equal to U, the energy density in the neighbourhood of the surface, will be exerted on it.[2] When a plane wave is incident at an angle, θ, on a completely absorbing surface a pressure equal to $U \cos^2\theta - U$ is now identical with the intensity of the incident wave since there is no reflexion—is exerted on it, and a tangential force equal to $U \sin\theta \cos\theta$ per unit area. The proof of these formulae may be left to the reader.

These consequences of electromagnetic theory have been

---

[1] This formula makes it clear that $\mathbf{p}$ is not always identical with the rate of flow of energy across the unit area. A group of waves travels with a velocity, v, which differs (see **§§ 9·3** and **27·1**) from u, the phase velocity, and the true rate of transport per unit area, namely vU, consequently differs from $\mathbf{p}$.

[2] Predicted by Maxwell (*Treatise*, § 792).

verified experimentally by Lebedew, Nichols and Hull, and Poynting.[1]

A very important case is that of the **black body radiation** or **full radiation,** which occupies a vacuous enclosure when the walls are maintained at some fixed temperature. This consists of electromagnetic waves of all wave-lengths and travelling in all directions. The amount of energy passing from one side to the other through any small area **dS** in the time dt and within the limits of the solid angle, $d\Omega(\equiv \sin\theta \; d\theta \; d\phi)$, may be written :

$$I \cos\theta \, \sin\theta \; d\theta \; d\phi \; dt \; \mathbf{dS}.[2]$$

The factor, I, is a constant, and consequently

$$\overline{\mathcal{E}_x{}^2} = \overline{\mathcal{E}_y{}^2} = \overline{\mathcal{E}_z{}^2} = \tfrac{1}{3}\overline{\mathcal{E}^2}$$

and

$$\overline{H_x{}^2} = \overline{H_y{}^2} = \overline{H_z{}^2} = \tfrac{1}{3} \; \overline{H^2}.$$

To get the pressure on the wall, imagine the X axis to be perpendicular to it. The pressure is the equal to $-t_{xx}$ and

$$- t_{xx} = \frac{1}{4\pi A} \{\tfrac{1}{6}\overline{K\mathcal{E}^2} + \tfrac{1}{6}\mu\overline{H^2}\}$$

or

$$p = \tfrac{1}{3}U. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad \mathbf{(27)}$$

It is therefore equal to one-third of the energy density.

§ 27·1. MASS AND ENERGY

The nearest approach that we can achieve to a plane harmonic wave (i.e. one in which the electric field intensity, for example, is expressed by

$$\mathcal{E}_x = \mathcal{E}_0 \cos\omega\{t - (lx + my + nz)/u\}$$

where $\mathcal{E}_0$ is a constant amplitude) is a superposition of an infinite number of such waves travelling in all directions within the limits of a very narrow solid angle containing the direction (l, m, n), and having an infinite number of frequencies (or periods) lying between ω and $\omega + d\omega$. Such a **group** of waves (as we shall term it) occupies a limited region, and includes all wave-lengths within narrow limits which we may represent by λ and $\lambda + d\lambda$. It can be shown that it will travel onwards without

---

[1] Nichols and Hull : *Phys. Review,* Vol. XIII, p. 293 (1901). Poynting : *Phil. Mag.,* Vol. IX, pp. 169, 475 (1905).

[2] Cf. the expression in **§ 12·1** for the number of molecules passing **dS** in the time dt.

sensible change in shape, and without spreading outwards appreciably, provided λ is very small compared with the dimensions of the space occupied by the group, or what amounts to the same thing, provided τ is very small compared with the time it requires to travel through its own length. Furthermore, it will, in this limiting case of very small τ, travel with a definite velocity, namely :

$$v = d(1/\tau)/d(1/\lambda), \quad . \quad . \quad . \quad . \quad (27\cdot1)$$

called the **group velocity** (cf. § **9·3**). The smaller τ is, in fact, the more nearly does the motion of the boundary of the group resemble that of the surface of a rigid body moving with a translatory velocity, v. Now we can, as we have seen, associate a definite momentum, **M** per unit volume, with the group ; and since we can also assign a definite velocity to it, we may regard it as having a mass, m, per unit volume, in accordance with the equation :

$$\mathbf{M} = mv. \quad . \quad . \quad . \quad . \quad (27\cdot11)$$

If we combine this with (**26·98**) we get the result

$$U = muv. \quad . \quad . \quad . \quad . \quad (27\cdot12)$$

Now when λ, or τ, is made very small indeed, the dispersion formula (**26·46**) becomes simply

$$n^2 = 1 + G\tau^2. \quad . \quad . \quad . \quad . \quad (27\cdot13)$$

where G is a constant.

Consequently

$$c^2/u^2 = 1 + G\tau^2,$$

or

$$c^2(1/\lambda^2) = G + 1/\tau^2.$$

Hence

$$c^2\frac{d(1/\lambda^2)}{d(1/\tau^2)} = 1,$$

or

$$c^2 = \frac{(1/\tau)d(1/\tau)}{1/\lambda\, d(1/\lambda)},$$

or, finally,

$$c^2 = uv. \quad . \quad . \quad . \quad . \quad . \quad (27\cdot14)$$

On substituting this result in (**27·12**) we find that

$$U = mc^2. \quad . \quad . \quad . \quad . \quad . \quad (27\cdot15)$$

This suggests that a determinate amount of energy is associated with a given mass—at all events with mass of electromagnetic origin. Electromagnetic energy is a more definite thing than is

energy in some of the forms in which we have met it. We have seen that energy has (as a consequence of the way in which it has been defined (§ 5·1)) an arbitrary constant associated with it. This arbitrariness is not, however, so serious as appears at first sight. In the case of electromagnetic energy we have fixed the arbitrary constant in such a way that the energy vanishes with the field. This is the natural procedure, since it is undesirable to ascribe properties to a thing which does not exist. We shall be led, in the study of relativity, to view the relationship

$$\text{Energy} = \text{mass} \times c^2 \quad . \quad . \quad . \quad . \quad \textbf{(27·16)}$$

as a universal one, if interpreted in the sense that wherever a mass exists it is associated with energy in accordance with **(27·16)**. It will be seen that it is only a question of a suitable choice of units to make **the measure of mass numerically identical with that of energy. We are thus led to regard mass as a form of energy.**

The adoption of equation **(27·16)** as a universal expression for the dependence of the mass of a body on its energy leads us to a contradiction, since equation **(5)** defines the mass of a body to be a constant characteristic of it and *independent of its velocity*. It is an instructive exercise to study this contradiction and to trace its origin. If we insist on the definition of mass in equation **(5)**—and possibly there is no other ground for abandoning it than that of convenience ; though this is a very reasonable ground—we throw back the responsibility for the contradiction on the meaning we assign to **force.**

The dynamical system of Hamilton, described in Vol. I, is based on measurements confined to bodies moving *very slowly* by comparison with the velocity of light (in free space). Now the only bodies which are moving, or can be caused to move, with velocities at all comparable with that of light are charged particles, α particles and electrons. Such particles moving in a magnetic field are subject to the force represented by **(22·6)**, a formula which has been derived from measurements made on systems with almost vanishingly small velocities. Let us consider the case of a charged particle subject to a force, due to a magnetic field only, and hence perpendicular to the direction of motion. In accordance with **(22·6)**, with a suitable choice of units ($a = 1$ and $\mu = 1$), we have apparently, when v is perpendicular to H :

$$Hev = \frac{mv^2}{r}, \quad . \quad . \quad . \quad . \quad \textbf{(27·17)}$$

where v is the velocity of the particle, r is the radius of the circle

in which it travels and m is the mass of the particle as defined by **(5)**. The right hand represents the product of the (presumed) constant, m, and the acceleration, v^2/r, of the particle ; while Hev is the force exerted on it. The product Hev will remain unchanged if the particle be subjected to a different field, αH, provided it moves with a velocity v/α, and equation **(27·17)** becomes

$$\alpha He\frac{v}{\alpha} = \frac{mv^2}{\alpha^2 r'}. \qquad\qquad . \quad . \quad . \quad . \quad . \mathbf{(27 \cdot 171)}$$

Newtonian dynamics thus requires :

$$\frac{mv^2}{\alpha^2 r'} = \frac{mv^2}{r},$$

or $\qquad\qquad\qquad \alpha^2 r' = r.$

Experiment indicates that this last equation does not hold. It turns out in fact that

$$\alpha^2 r' > r$$

when

$$\alpha < 1.$$

We have two alternatives before us : we may decide to retain **(5)**, i.e. to continue to define the force on a particle as the product of a constant (characteristic of the particle) and its acceleration, in which case the expression **(22·6)** will represent the force correctly only in the limiting case of small velocities ; or we may retain **(22·6)** as expressing the force, and modify **(5)**. The latter course is the one which has been adopted. That is to say, we define force so that the formula **(22·62)** will always hold, whatever may be the velocity of the charged particle. We maintain the formulae **(5·01)** and **(5·011)**, so that

$$\mathbf{F} = \frac{d}{dt}(m\mathbf{v}), \qquad . \quad . \quad . \quad . \quad . \mathbf{(27 \cdot 18)}$$

in which m is no longer restricted to be a constant ; but must asymptotically approach the m of **(5)** as the velocity of the particle approaches the limit zero.

§ 27·2. Dependence of Mass on Velocity

Let $\mathbf{F}$ be the force exerted on a particle. In accordance with **(27·18)**

$$F_x = \frac{d}{dt}(mv_x).$$

Consequently

$$F_x = m\frac{dv_x}{dt} + v_x\frac{dm}{dt},$$

or

$$F_x = m\frac{dv_x}{dt} + v_x\frac{dm}{dv}\cdot\frac{dv}{dt}. \quad . \quad . \quad . \quad . \quad \textbf{(27·2)}$$

There are two interesting special cases : (1) that in which the force is normal to the velocity, i.e. $v_x = 0$, and (2) that in which the force has the same direction as the velocity. In case (1) **(27·2)** becomes

$$F_x = m\frac{dv_x}{dt}, \quad . \quad . \quad . \quad . \quad \textbf{(27·21)}$$

so that m is the factor with which the acceleration must be multiplied to give the force, when the latter is perpendicular to the velocity.[1] In case (2) $v_x = \mathbf{v}$ and

$$\mathbf{F} = \left(m + \mathbf{v}\frac{dm}{d\mathbf{v}}\right)\frac{d\mathbf{v}}{dt}. \quad . \quad . \quad . \quad \textbf{(27·22)}$$

In consequence of the relation **(27·21)** m was formerly described as the **transverse mass**, whereas

$$m' = m + \mathbf{v}\frac{dm}{d\mathbf{v}} \quad . \quad . \quad . \quad . \quad \textbf{(27·221)}$$

(27·22) was described as the **longitudinal mass.**

The work done on the particle during a small displacement **dl** in the direction of the force, i.e. under the condition (2), is

$$\mathbf{Fdl} = \left(m + \mathbf{v}\frac{dm}{d\mathbf{v}}\right)\frac{d\mathbf{v}}{dt}\mathbf{dl},$$

or, since

$$\frac{\mathbf{dl}}{dt} = \mathbf{v},$$

$$\mathbf{Fdl} = \left(m + \mathbf{v}\frac{dm}{d\mathbf{v}}\right)vdv. \quad . \quad . \quad . \quad \textbf{(27·23)}$$

This will represent the increase in the energy of the particle. Let us, adopting **(27·16)**, equate it to c^2dm ; thus

$$c^2dm = \left(m + \mathbf{v}\frac{dm}{d\mathbf{v}}\right)vdv.$$

---

[1] This statement about mass should be kept in mind in studying the experimental investigation of the variation of mass with velocity.

Therefore

$$(c^2 - v^2)dm = mvdv \; ;$$

which gives (cf. § 9·4)

$$m = m_0\left(1 - \frac{v^2}{c^2}\right)^{-1/2}, \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad (27·24)$$

an equation which we shall usually write in the abbreviated form :

$$m = m_0\gamma. \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad (27·241)$$

in which γ is used to represent $(1 - v^2/c^2)^{-1/2}$.

On substituting (27·241) in (27·221) we find for the longitudinal mass the expression :

$$m' = m_0\gamma^3. \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad (27·242)$$

It has turned out to be without any theoretical or practical importance. The kinetic energy of the particle must be equal to

$$(m - m_0)c^2, \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad (27·25)$$

since it is the difference between its energy when in motion and when at rest. On substituting in this formula the expression (27·24) for m, we find that it reduces, for small velocities, to $m_0v^2/2$, as we should expect.

§ 27·3. Bucherer's Experiment

A very elegant method of investigating experimentally the dependence of mass on velocity was devised by Bucherer.[1] His apparatus, illustrated in Fig. 27·3 (plan) and Fig. 27·31 (a) (elevation), consisted of a condenser, A, made of two parallel circular plates of like dimensions and separated by a minute distance. The condenser was situated in a vessel (exhausted during the experiment) with a cylindrical wall, B, the axis, O, of which was coincident with that of the condenser plates.
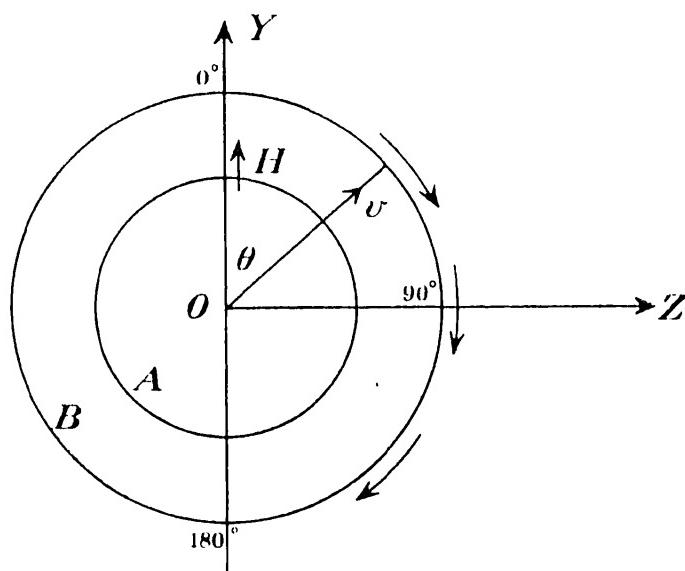
Fig. 27·3

[1] Bucherer : *Ann. d. Phys.*, IV, Vol. XXVIII, p. 513 (1909).

We shall suppose this axis to coincide with the X axis of the co-ordinates we are going to use in the description of the experiment. The Y and Z axes we shall place in the plane of the narrow interspace between the condenser plates, which in Fig. 27·3 is in the plane of the paper. A uniform electric field in the direction of the X axis, or in the opposite one, is produced between the condenser plates by connecting them to the poles of a suitable battery. This field drops to zero at the edge of the condenser and may be regarded as non-existent in the region outside the condenser. A uniform magnetic field is generated in the Y direction, or in the opposite direction (i.e. parallel to the condenser plates), by means of a current which flows round the windings of a suitable coil or coils outside the apparatus, and the axis of which is parallel to the condenser plates. A photographic film was extended round the interior surface of the cylindrical wall, B, of the containing vessel. The part of this film extending from 0° to 180° over the part of

Fig. 27·31

the cylindrical wall marked in Fig. 27·3 by arrows is shown in Fig. 27·31 (b).[1] The median line, $\alpha\alpha$, is in the plane of the narrow interspace between the condenser plates. A speck of radium fluoride was placed at the centre, O, (Fig. 27·3 and Fig. 27·31 (a)) of the interspace. This emitted, among other things, β particles (electrons) of various velocities ranging almost to the velocity of light. So much for the description of the apparatus.

The electrons or β particles projected from the radioactive material in directions parallel to the surfaces of the condenser plates will, in the absence of applied fields, strike the film and produce the line $\alpha\alpha$ (Fig. 27·31 (b)). In the presence of the electric and magnetic fields, the X component of the force on a β particle is by (22·63)

$$F_x = e\left\{ \mathcal{E}_x + \frac{1}{c}(v_y H_z - v_z H_y) \right\} ;$$

and when, as we are supposing, the electric field is in the X

---

[1] No attempt has been made to construct the two diagrams, (a) and (b), on the same scale.
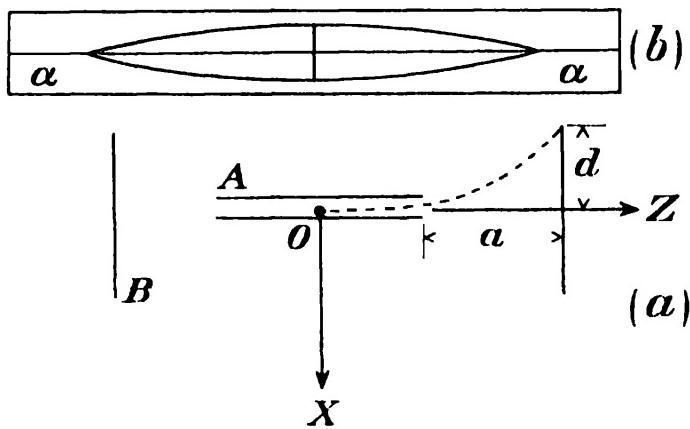
direction and the magnetic field in the Y direction, this becomes

$$F_x = e\left(\mathcal{E} - \frac{v_z H}{c}\right),$$

in which we have replaced $\mathcal{E}_r$ by $\mathcal{E}$ and H_y by H. In order that the particle may emerge F_r must be zero and v must be in a plane parallel to the surfaces of the condenser plates. Thus v_z is equal to $v \sin \theta$. Therefore

$$\mathcal{E} = \frac{v \sin \theta . H}{c}.$$

or

$$\frac{v}{c} = \frac{\mathcal{E}}{H \sin \theta}. \qquad . \qquad . \qquad . \qquad . \qquad (27 \cdot 3)$$

After leaving the interspace between the condenser plates the particle is subject to the influence of the magnetic field only and strikes the film above or below the line $\alpha\alpha$. Which of these two possibilities occurs depends on the sign of e, which is negative for β particles, and on whether H is in the Y direction or the opposite one. In the experiment both directions for the two fields were used; so that traces appeared on the film above and below the horizontal line $\alpha\alpha$. These are illustrated for the case $\mathcal{E}/H = 1/2$ in Fig. 27·31 (b). Bucherer found that in the latter case they joined the line $\alpha\alpha$ at points where θ had the values 30° and 150°; for both of which $\sin \theta$ is 1/2. The immediate inference is that β particles do not emerge in directions for which θ is less than 30° or more than 150°, or, if they do, they are undeviated by the magnetic field. Assuming for the moment that some β particles really do reach portions of the film where $\theta < 30°$ or $\theta > 150°$, this can only occur because the force due to the field does not cause them to deviate from the path they would follow in the absence of the field, and we may infer that these particles follow rectilinear paths. Hence r is infinite, and consequently so is m, since mv^2/r, the absolute value of the force exerted on a particle, is finite. We conclude therefore that the β particles are ejected within the angular range from 30° to 150° and that c is the upper limit of their velocities, in accordance with (27·3), $\mathcal{E}/H$ being 1/2.

Bucherer verified the theoretical law already given for the dependence of the mass of a particle on its velocity by measuring the distance, d, (corresponding to various velocities) of the bow-shaped lines from the line $\alpha\alpha$ at the centre, i.e. where $\theta = 90°$. The β particles ejected under this angle move in the vertical plane at right angles to H (the plane of the paper in Fig.

27·31 (a)). Outside the condenser each one travels in a circle the radius of which is

$$r = (a^2 + d^2)/2d,$$

in which a and d have the meanings indicated in the figure. Hence the equation of motion of a particle is

$$\frac{Hev}{c} = \frac{mv^2}{r},$$

or

$$\frac{He}{c} = \frac{2mvd}{a^2 + d^2}, \quad . \quad . \quad . \quad . \quad (27\text{·}31)$$

and the corresponding velocity is

$$v = c\mathcal{E}/H. \quad . \quad . \quad . \quad . \quad (27\text{·}32)$$

The measurement of $\mathcal{E}$, H, a and d enabled Bucherer to evaluate m/e and the corresponding velocity by the use of (27·31) and (27·32) or equivalent formulae. For a given value of θ, in this case 90°, and a given value of the ratio $\mathcal{E}/H$, only β particles with the velocity expressed by (27·32) can emerge and hit the film. Experiments of other kinds have demonstrated the invariability of the charge e on the β particle, so that the law expressing the dependence of m/e on the velocity is identical with that for m itself. Bucherer's measurements verified the law already suggested, which is due to H. A. Lorentz, and he found for e/m_0 (where m_0 is the limiting value for $v = 0$) the value $1\text{·}77 \times 10^{-7}$ ordinary E.M. units per gram.

§ 27·4. Second Order Equations of the Electromagnetic Field

We shall adopt the field equations (26·93) which are eminently suitable for the purposes of the electron theory ; but we shall write them in the following form :

$$\left.\begin{array}{c} \dfrac{K}{a}\dfrac{\partial\mathcal{E}}{\partial t} + \dfrac{4\pi A \rho \mathbf{v}}{a} = \mathbf{curl\ H,} \\[2mm] \mathbf{div\ \mathcal{E}} = \dfrac{4\pi A}{K}\rho, \\[2mm] -\dfrac{\mu}{a}\dfrac{\partial \mathbf{H}}{\partial t} = \mathbf{curl\ \mathcal{E},} \\[2mm] \mathbf{div\ H} = 0. \end{array}\right\} \quad . \quad . \quad . \quad (27\text{·}4)$$

In this form the equations are adapted for any system of units whatever. The choice of units is decided by the values assigned

to the numerical constants A and a and, as we have seen, by the unit adopted for K or for μ. We are assuming that K and μ are constants. Indeed this is implied by the form of the equations, since the displacement current density, for example, is $\partial D/\partial t$ and may only be written $K\partial\mathcal{E}/4\pi A\partial t$ when K is a constant. Similarly, the electrical divergence equation is really $\mathbf{div\ D} = \rho$, and only takes the form given in (27·4) when K is a constant. Similar observations may be made about the equations containing μ.

Let us now introduce a vector potential, $\mathbf{A}$, so defined that

$$\mathbf{D}_m = \mathbf{curl\ A}. \quad . \quad . \quad . \quad . \quad \textbf{(27·41)}$$

This equation does not of course completely define $\mathbf{A}$, since if a vector $\mathbf{A}$ has the property (27·41), all those vectors which differ from it by vectors which are gradients have the same property (cf. § 23). To complete the definition of $\mathbf{A}$ we must, as equation (23·04) indicates, specify what its divergence is to be. This we shall do later. Meanwhile we shall substitute (27·41) in the first equation (27·4). We thus get :

$$\nabla^2\mathbf{A} + \frac{\mu K}{4\pi Aa}\frac{\partial\mathcal{E}}{\partial t} = -\frac{\mu\rho\mathbf{v}}{a} + \mathbf{grad\ div\ A.}^1 \quad . \quad \textbf{(27·42)}$$

On substituting (27·41) in the third equation (27·4) we get :

$$\mathbf{curl}\left(-\frac{4\pi A}{a}\frac{\partial\mathbf{A}}{\partial t}\right) = \mathbf{curl\ }\mathcal{E},$$

and consequently $\mathcal{E}$ and $-4\pi A\ \partial\mathbf{A}/a\partial t$ differ by a vector which is a gradient. We shall write this gradient in the form : $\mathbf{grad}\ (-V)$, where V is some scalar quantity. Consequently

$$\mathcal{E} = -\mathbf{grad}\ V - \frac{4\pi A}{a}\frac{\partial\mathbf{A}}{\partial t}. \quad . \quad . \quad . \quad \textbf{(27·43)}$$

In a static field $\partial\mathbf{A}/\partial t$ is of course equal to zero, and $\mathcal{E}$ becomes $-\mathbf{grad}\ V$. Hence V is in this case the electrostatic potential. We shall give it the general name : **scalar potential**.

If we substitute the expression (27·43) for $\mathcal{E}$ in (27·42) we get

$$\nabla^2\mathbf{A} - \frac{\mu K}{a^2}\frac{\partial^2\mathbf{A}}{\partial t^2} = -\frac{\mu\rho\mathbf{v}}{a} + \mathbf{grad}\left\{\mathbf{div\ A} + \frac{\mu K}{4\pi Aa}\frac{\partial V}{\partial t}\right\}. \quad \textbf{(27·44)}$$

We now substitute (27·43) in the second of the equations (27·4) and so obtain

$$\mathbf{div}\left\{-\mathbf{grad}\ V - \frac{4\pi A}{a}\frac{\partial\mathbf{A}}{\partial t}\right\} = \frac{4\pi A}{K}\rho,$$

---

^{1} Cf. equation (2·45).

or

$$\nabla^2 V = -\frac{4\pi A}{K}\rho - \frac{4\pi A}{a}\frac{\partial}{\partial t}(\text{div A}).$$

To both sides of this latter equation we may add the expression :

$$-\mu K \partial^2 V/a^2\partial t^2,$$

and so obtain :

$$\nabla^2 V - \frac{\mu K}{a^2}\frac{\partial^2 V}{\partial t^2} = -\frac{4\pi A\rho}{K} - \frac{4\pi A}{a}\frac{\partial}{\partial t}\left\{\text{div A} + \frac{\mu K}{4\pi Aa}\frac{\partial V}{\partial t}\right\}. \quad . \quad (27\cdot45)$$

We can bring about a desirable uniformization of equations (27·44) and (27·45) by introducing two new symbols A_w and w with the meanings :

$$\left.\begin{array}{l} A_w \equiv i\mu^{1/2}K^{1/2}V/4\pi A, \\ dw \equiv iadt/\mu^{1/2}K^{1/2}. \end{array}\right\} \quad . \quad . \quad (27\cdot451)$$

The expression in the brackets { } then takes the form :

$$(\Box, \text{A}) \equiv \frac{\partial A_x}{\partial x} + \frac{\partial A_y}{\partial y} + \frac{\partial A_z}{\partial z} + \frac{\partial A_w}{\partial w}, \quad . \quad (27\cdot452)$$

which is a sort of 4-dimensional divergence. We may now combine equations (27·44) and (27·45) in the single statement :

$$\Box^2 \text{A} = -\text{s} + \Box (\Box, \text{A}), \quad . \quad . \quad (27\cdot46)$$

in which A now includes the 4 components A_x, A_y, A_z and A_w. We may call it a vector in an extended sense (cf. § 10·5). The extended vector s has the components $\frac{\mu\rho v_x}{a}$, $\frac{\mu\rho v_y}{a}$, $\frac{\mu\rho v_z}{a}$, $\frac{\mu\rho v_w}{a}$, where v_w means $ia/K^{1/2}\mu^{1/2}$.

We have still to complete the definition of the extended vector **A**. We shall do this in a way exactly analogous to that used in § 23·1. We shall choose that vector **A** the extended divergence of which, namely $(\Box, \text{A})$, is equal to zero. There is such a vector **A**, as we shall see. We thus have finally

$$\Box^2\text{A} = -\text{s} = -\frac{\mu\rho\text{V}}{a}, \quad . \quad . \quad . \quad (27\cdot47)$$

which we may call an **extended Poisson's equation.**

The case of special interest to us is that in which the dielectric medium is the aether. The phase velocity, u, then becomes the velocity represented by c, and the equations (27·451) become, if we employ Lorentz-Heaviside units :

$$\left.\begin{array}{l} A_w = iV, \\ dw = icdt. \end{array}\right\} \quad . \quad . \quad . \quad . \quad (27\cdot48)$$

The fourth component of the extended Poisson's equation **(27·47)** now becomes :

$$\nabla^2 V - \frac{1}{c^2}\frac{\partial^2 V}{\partial t^2} = -\rho, \quad . \quad . \quad . \quad \textbf{(27·49)}$$

and, of the remaining three, a typical one is

$$\nabla^2 A_x - \frac{1}{c^2}\frac{\partial^2 A_x}{\partial t^2} = -\frac{\rho v_x}{c}. \quad . \quad . \quad \textbf{(27·491)}$$

The expression on the left hand of **(27·49)** is termed the **lorentzian** of V (**lor** V) and we may write **lor A** instead of $\square^2\textbf{A}$.

§ 27·5. Solution of the Extended Poisson's Equation

The equation **(27·49)** reduces to the simple form of Poisson's equation when the time differential quotient in it is zero. If instead of this the term on the right vanishes, it becomes the familiar equation of wave propagation. These facts suggest a solution of the equation. We are in fact led to anticipate a solution resembling **(18·85)**, namely :

$$V = \frac{1}{4\pi}\int\int\int \frac{[\rho]}{r}dx\,dy\,dz, \quad . \quad . \quad . \quad \textbf{(27·5)}$$

in which V is the scalar potential at some point x_0, y_0, z_0, conveniently the origin of co-ordinates, at an instant t, and where the volume integral is extended over a region enclosed by a surface—we are ignoring for the moment the possibility of a contribution represented by an integral extended over the surface. We naturally represent the contribution of a volume element $dx\,dy\,dz$ as due to the charge density, $[\rho]$, existing there, not at the instant, t, but at the earlier instant, $t - r/c$, since its influence is propagated with the velocity c. It is convenient to use the following notation :
If R be a function

$$R = f(x,\,y,\,z,\,t),$$

then $[R]$ shall have the meaning :

$$[R] = f(x,\,y,\,z,\,\tau), \quad . \quad . \quad . \quad \textbf{(27·51)}$$

where

$$\tau = t - r/c. \quad . \quad . \quad . \quad \textbf{(27·511)}$$

and

$$r = \sqrt{(x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2}\,|$$

or, if $(x_0,\,y_0,\,z_0)$ is the origin,

$$r = \sqrt{x^2 + y^2 + z^2}\,|\,.$$

The symbols $\partial/\partial x$, $\partial/\partial y$, $\partial/\partial z$ and $\partial/\partial t$ will be used for that partial differentiation in which x, y, z, and t are the independent variables. A differentiation in the direction of the vector, $\mathbf{r} \equiv (x, y, z)$, we shall represent by the symbol δ. It is easy to prove that

$$\frac{\partial[R]}{\partial x} = \left[\frac{\partial R}{\partial x}\right] - \frac{x}{rc}\left[\frac{\partial R}{\partial t}\right]. \quad \cdots \quad \textbf{(27·52)}$$

If we use $\partial/\partial\tau$ to represent a partial differentiation of a function $f(x, y, z, \tau)$ with respect to τ, when the explicitly appearing variables x, y and z are kept constant, then

$$\frac{\partial[R]}{\partial\tau} = \left[\frac{\partial R}{\partial t}\right]. \quad \cdots \quad \textbf{(27·53)}$$

Further, in consequence of the definition of δ,

$$\frac{\delta[R]}{\delta r} = \frac{x}{r}\frac{\partial[R]}{\partial x} + \frac{y}{r}\frac{\partial[R]}{\partial y} + \frac{z}{r}\frac{\partial[R]}{\partial z},$$

and therefore by **(27·52)**

$$\frac{\delta[R]}{\delta r} + \frac{1}{c}\frac{\partial[R]}{\partial\tau} = \frac{x}{r}\left[\frac{\partial R}{\partial x}\right] + \frac{y}{r}\left[\frac{\partial R}{\partial y}\right] + \frac{z}{r}\left[\frac{\partial R}{\partial z}\right]. \quad \cdots \quad \textbf{(27·54)}$$

To derive the solution required, we start out from the expression :

$$\mathbf{div}\left\{\frac{1}{r}[\mathbf{grad}\ V]\right\}.$$

We have

$$\frac{\partial}{\partial x}\left\{\frac{1}{r}\left[\frac{\partial V}{\partial x}\right]\right\} = \frac{1}{r}\frac{\partial}{\partial x}\left[\frac{\partial V}{\partial x}\right] - \frac{x}{r^3}\left[\frac{\partial V}{\partial x}\right]$$

and two similar equations. By making use of **(27·52)** we get

$$\frac{\partial}{\partial x}\left\{\frac{1}{r}\left[\frac{\partial V}{\partial x}\right]\right\} = \frac{1}{r}\left[\frac{\partial^2 V}{\partial x^2}\right] - \frac{x}{r^2 c}\left[\frac{\partial^2 V}{\partial x \partial t}\right] - \frac{x}{r^3}\left[\frac{\partial V}{\partial x}\right]$$

and two further equations similarly associated with the Y and Z axes. When we add all three together and remember that

$$\frac{1}{r}\left[\frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} + \frac{\partial^2 V}{\partial z^2}\right] = \frac{-[\rho]}{r} + \frac{\dfrac{1}{c^2}\left[\dfrac{\partial^2 V}{\partial t^2}\right]}{r},$$

we obtain

$$\mathbf{div}\left\{\frac{1}{r}[\mathbf{grad}\ V]\right\} = \frac{-[\rho]}{r} + \frac{\dfrac{1}{c^2}\left[\dfrac{\partial^2 V}{\partial t^2}\right]}{r} - \frac{1}{rc}\left\{\frac{\delta}{\delta r}\left[\frac{\partial V}{\partial t}\right] + \frac{1}{c}\left[\frac{\partial^2 V}{\partial t^2}\right]\right\}$$

$$- \frac{1}{r^2}\left\{\frac{\delta[V]}{\delta r} + \frac{1}{c}\left[\frac{\partial V}{\partial t}\right]\right\}.$$

$$= \frac{-[\rho]}{r} - \frac{1}{r^2}\left(\frac{\delta}{\delta r}\left\{\frac{r}{c}\left[\frac{\partial V}{\partial t}\right]\right\} + \frac{\delta[V]}{\delta r}\right), \quad \text{by } \textbf{(27·54)}.$$

Now multiply both sides by the volume element $dx\,dy\,dz$ and integrate (as in **§ 3·1**) over a region enclosed by some given surface (within which the origin is situated) and by the surface of a sphere of small radius, R, with the origin as centre. We thus obtain, after applying the theorem of Gauss to the left-hand side :

$$\iint \frac{1}{r}\left[\frac{\partial V}{\partial n}\right]dS = -\iiint \frac{[\rho]}{r}dx\,dy\,dz - \iiint \frac{\delta}{\delta r}\left\{\frac{r}{c}\left[\frac{\partial V}{\partial t}\right]+[V]\right\}\delta r\,d\Omega.$$

The volume element in the last integral is $r^2 d\Omega\delta r$, where $d\Omega$ is the elementary solid angle with the axis r. Integration with respect to δr gives

$$-\iint \left\{\frac{r}{c}\left[\frac{\partial V}{\partial r}\right]+[V]\right\}d\Omega$$

which, over the outer surface, yields

$$-\iint \left\{\frac{1}{rc}\left[\frac{\partial V}{\partial t}\right]+\frac{1}{r^2}[V]\right\}\cos\theta\,dS$$

and over the inner surface (sphere of radius R) when R approaches the limit zero :

$$+4\pi V_{0,\ 0,\ 0,\ t}.$$

Finally therefore we have, if we replace

$$\cos\theta \text{ by } \partial r/\partial n$$

and

$$\cos\theta/r^2 \text{ by } -\partial(1/r)/\partial n,$$

$$V_{0,\ 0,\ 0;\ t} = \frac{1}{4\pi}\iiint \frac{[\rho]}{r}dx\,dy\,dz$$
$$+\frac{1}{4\pi}\iint \left\{\frac{1}{r}\left[\frac{\partial V}{\partial n}\right] - [V]\frac{\partial\left(\frac{1}{r}\right)}{\partial n} + \frac{1}{rc}\left[\frac{\partial V}{\partial t}\right]\frac{\partial r}{\partial n}\right\}dS. \tag{27·55}$$

We have of course analogous expressions for A_x, A_y and A_z, which need not be written down here. For A_x of **(27·491)**, for example, we replace $[\rho]$ in **(27·55)** by $\left[\dfrac{\rho v_x}{c}\right]$ and make corresponding changes in the surface integral.

The solution **(27·55)** is usually called **Kirchhoff's solution.** The important special cases are (a) that in which the surface integral vanishes, and which we shall use in **§ 27·6**, and (b) that

in which the volume integral vanishes. We shall apply the latter (§ 31·8) to obtain an expression for **Huygens' principle.**

It is easy to show (by the use of the theorems of § 23 and the equation of conservation (22·751), that the components of the extended vector **A**, as given by Kirchhoff's solution, satisfy

$$(\square, \mathbf{A}) = 0.$$

If this were not the case our solution would have no value since the validity of (27·47) depends on it.

§ 27·6.   Field of a Moving Point Charge

We shall now derive, with the aid of (27·55), an expression for the scalar potential at the point (0, 0, 0) at an instant, t, due to a point charge, i.e. to a small charged body. The term ' small ' means, of course, small compared with r, its distance from (0, 0, 0). Let us consider first a cylindrical volume element of the small charged body, dl in length and dA in cross-sectional area, and let us imagine it to be moving with the velocity, v, along the line r, as indicated in Fig. 27·6, the axis, dl, of the element being parallel to r. We may place the surface. over which the surface integral of (27·55) is extended, so far away that the

Fig. 27·6

integral is zero (cf. § 3·1) in which case (27·55) reduces to the volume integral only. Let us suppose the end of the element next to (0, 0, 0) is distant r at the instant $t - r/c$. The density, ρ, at that end of the element is contributing to V at this particular instant, while at this instant the other end is dl farther away. It has contributed at some earlier instant, when it was in the position indicated by the broken line. This earlier instant may be denoted by $t - (r + dr)/c$. The interval of time between the two instants is dr/c, and during this interval the element has travelled the distance vdr/c. Hence

$$dl - dr = vdr/c,$$

and so $$dr = dl/(1 + v/c).$$

Now it is obvious that within the volume $dr\, dA$,

$$[\rho] = \rho.$$

While everywhere else $[\rho] = 0.$
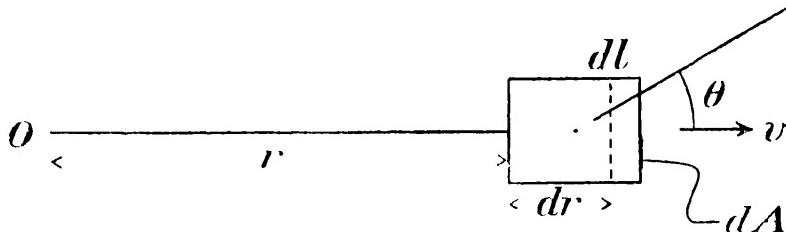
The difference therefore between the electrostatic problem and the present one, is that, while in the former the potential at $(0, 0, 0)$ will be $\rho \, dl \, dA / r$, in the latter it will be

$$\rho \, dr \, dA / r.$$

If e be the charge in the element

$$e = \rho \, dl \, dA,$$

consequently by **(27·55)**,

$$V_{0, 0, 0, t} = \frac{e}{4\pi r (1 + v/c)}, \quad \cdot \quad \cdot \quad \cdot \quad \textbf{(27·6)}$$

and this formula will obviously hold for the whole of the small body, provided e means the whole charge.

If we go a step further and suppose the velocity, v, to make an angle, θ, with the direction of r (Fig. 27·6), then obviously **(27·6)** must be amended to

$$V_{0, 0, 0, t} = \frac{e}{4\pi r (1 + v \cos \theta / c)}. \quad \cdot \quad \cdot \quad \textbf{(27·61)}$$

This formula is of course expressed in terms of Lorentz-Heaviside units and refers to the special case (empty space) for which $\mu = K = 1$.

The appropriate expressions for the components A_x, A_y and A_z of the vector potential can be obtained at once by inspection. In calculating A_x, for example, we have to replace ρ in **(27·55)** by $\rho v_x / c$, and so we get :

$$A_x = \frac{e \, v_x}{4\pi r c (1 + v \cos \theta / c)}, \quad \cdot \quad \cdot \quad \textbf{(27·62)}$$

and analogous expressions for the remaining components.

These formulae for the scalar and vector potentials associated with a moving particle enable us, with the aid of **(27·41)** and **(27·43)**, to find out all about the field of the moving charged particle.

§ 27·7. Radiation of Energy from an Accelerated Charged Particle

We imagine a spherical surface of radius r described round the particle as centre and proceed to find the amount of energy passing through it in the unit time. To accomplish this we must evaluate Poynting's vector at each point on the spherical surface, and then integrate over the surface. This means that we must

find the vectors $\mathcal{E}$ and H at each point on the spherical surface, and for this purpose we have to make use of the formulae of § 27·6 for the electromagnetic potentials. We shall be content with a result which is true for small velocities,[1] and hence (27·61) and (27·62) simplify to

$$V = \frac{e}{4\pi r},$$

and

$$A_x = \frac{ev_x}{4\pi cr}$$ . . . . . (27·7)

respectively.

The particle is situated at O (Fig. 27·7) at some instant, t, and we may choose co-ordinates in which it is, at that instant,
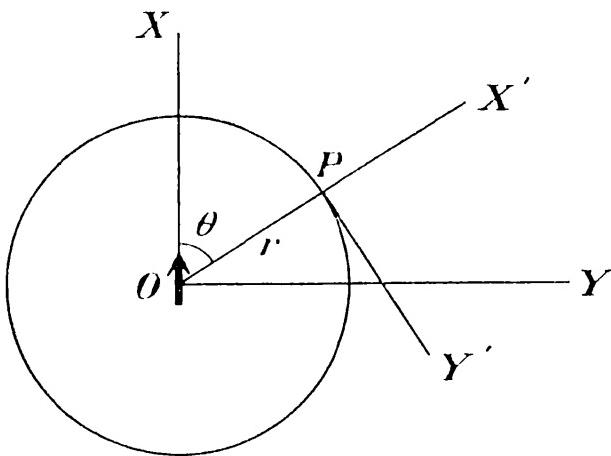
Fig. 27·7

at rest, and the X axis of which has the direction of the acceleration, **a**. To calculate the amount of energy crossing the spherical surface per unit area per unit time, all we require is the component of Poynting's vector, in an outward direction, normal to the surface. Let us place the origin of a provisional set of co-ordinate axes, X', Y' and Z' at a point, P, in the XY plane, as shown in the figure, the X' axis having the direction of r, i.e. of the outward normal to the spherical surface, and Z' being parallel to Z. Calling this component simply p, we have

$$p = c\{\mathcal{E}_{y'}'H_{z}' - \mathcal{E}_{z}'H_{y'}'\},$$

or

$$p = c\,\mathcal{E}_{y'}'H_{z}',$$ . . . . . (27·71)

since all the components of H vanish except H_z.

Now, by (27·43), $\mathcal{E}_{y'}'$ is made up of two contributions. One, the contribution of the scalar potential, is obviously zero ; since

[1] For a more elaborate investigation see O. W. Richardson's *Electron Theory of Matter*, 2nd edition, Chapter XII.

the part of the field intensity due to it is directed along X'. The component $\mathcal{E}_y'$ is therefore due solely to the time variation of the vector potential. Now we have placed our co-ordinates so that $\mathbf{A}$ is parallel to X. Its contribution to $\mathcal{E}$ is

$$- \frac{1}{c} \frac{\partial}{\partial t} \left(\frac{ev}{4\pi cr} \right),$$

therefore

$$\mathcal{E}_y' = \frac{1}{c} \frac{\partial}{\partial t} \left(\frac{ev}{4\pi cr} \right) \sin \theta,$$

or

$$\mathcal{E}_y' = \frac{e\mathbf{a}}{4\pi c^2 r} \sin \theta, \quad . \quad . \quad . \quad . \quad \textbf{(27·72)}$$

where $\mathbf{a}$ is the acceleration of the particle. Since $\mu = 1$, we have further

$$H_z' = \frac{\partial A_y}{\partial x'} - \frac{\partial A_x}{\partial y'},$$

or

$$H_z' = - \frac{\partial \mathbf{A}}{\partial y'} = - \frac{\partial}{\partial y'} \left(\frac{ev}{4\pi rc} \right) = \frac{ev}{4\pi r^2 c} \frac{\partial r}{\partial y'} - \frac{e}{4\pi rc} \frac{\partial v}{\partial y'}.$$

Therefore

$$H_z' = \frac{ev}{4\pi r^2 c} \frac{\partial r}{\partial y'} - \frac{e}{4\pi rc} \frac{\partial v}{\partial \xi} \frac{\partial \xi}{\partial y'},$$

where ξ means $t - r/c$. This becomes finally

$$H_z' = \left(\frac{ev}{4\pi r^2 c} + \frac{e\mathbf{a}}{4\pi rc^2} \right) \sin \theta,$$

since $\partial v/\partial \xi = \partial v/\partial t = \mathbf{a}$ and $\partial \xi/\partial y' = \partial(t - r/c)/\partial y' = - \sin \theta/c$, and we may conveniently write it in the form :

$$H_z' = \frac{e}{4\pi rc} \left\{ \frac{v}{r} + \frac{\mathbf{a}}{c} \right\} \sin \theta, \quad . \quad . \quad . \quad \textbf{(27·73)}$$

hence the normal component of Poynting's vector is (cf. **27·71**) :

$$p = \frac{e^2 \mathbf{a}}{16\pi^2 r^2 c^2} \left\{ \frac{v}{r} + \frac{\mathbf{a}}{c} \right\} \sin^2 \theta. \quad . \quad . \quad \textbf{(27·74)}$$

This expression has been deduced for a point, P, in the plane XY; but the symmetry of the problem assures us that it is

valid for any point on the spherical surface. The energy flowing outwards through the latter per unit time is

$$\int_0^\pi \int_0^{2\pi} p\, r^2 \sin\theta\, d\theta\, d\phi.$$

Substituting **(27·74)** for p in this integral and carrying out the integration we get :

$$\boxed{\frac{e^2 a^2}{6\pi c^3} + \frac{e^2 a v}{6\pi c^2 r}} \qquad . \qquad . \qquad . \qquad . \qquad \textbf{(27·75)}$$

The energy in the field round about a slowly moving charged particle is, according to **(26·13)**,

$$\frac{e^2 v^2}{12\pi c^2 r} \qquad . \qquad . \qquad . \qquad . \qquad . \qquad \textbf{(27·76)}$$

outside a sphere of radius r, since $\mu = 1$, and we are now using units for which $A = 1/4\pi$, and the a of **(26·13)** is here c. This is the magnetic energy which travels along with the particle. When the particle is accelerated there will be (in consequence of **(27·76)**) a flow of energy through the sphere of radius r equal to

$$\frac{d}{dt}\left(\frac{e^2 v^2}{12\pi c^2 r}\right)$$

or

$$\frac{e^2 v a}{6\pi c^2 r},$$

since a means the acceleration dv/dt. This term in **(27·75)** represents therefore the energy flow required to build up the magnetic field appropriate to the velocity of the moving charge, and cannot be regarded as energy *leaving it*. Indeed, the greater the radius, r, of the spherical surface, the smaller is the rate at which energy flows outwards through it. This part of the energy flow is appropriately termed by Langevin the **wave of re-organization**. The remaining term, namely

$$e^2 a^2 / 6\pi c^3, \qquad . \qquad . \qquad . \qquad . \qquad . \qquad \textbf{(27·77)}$$

represents (since it does not contain r) the energy passing through *any surface* surrounding the charged particle. It is the energy which *really leaves it* in the unit time. The expression **(27·77)** was first derived by Larmor and we shall term it **Larmor's formula.**

A particle with a charge, e (in Lorentz-Heaviside units), and

in simple harmonic vibration of linear amplitude A, radiates energy at the average rate :

$$\frac{e^2 A^2 \omega^4}{12\pi c^3} \quad \text{or} \quad \frac{4\pi^3 e^2 A^2 \nu^4}{3c^3},$$

where ω is the angular frequency, and ν is the frequency in the sense of the number of vibrations per unit time. The proof of these formulae is left to the reader.

§ 27·8. The Scattering of Radiation

A charged particle in the path of a beam of electromagnetic waves will scatter the energy in the beam, because there will in general be a force exerted on it which will give it an acceleration. Let the particle be situated at the origin of rectangular co-ordinates (Fig. 27·8), and suppose in the first instance plane waves travelling in the Z direction, and polarized so that the electric field intensity is in the line of the X axis. We may, for the reasons explained in § 26·4, ignore the contribution of the magnetic field of the waves to the force exerted on the particle, and take this consequently to be due solely to the electric field intensity. The acceleration of the particle is therefore along the line of the X axis and (27·74) may be applied here. We shall drop the term in (27·74) containing v, since, as we have seen, it does not contribute to the energy radiated by the particle. The formula therefore simplifies to
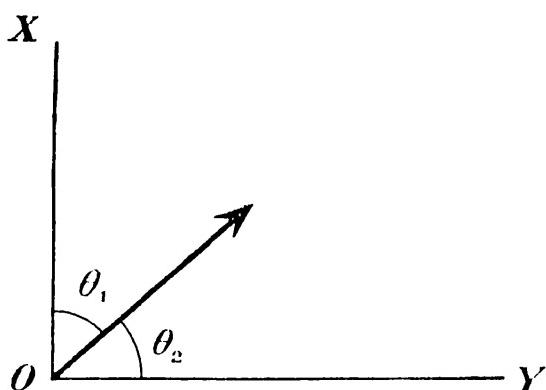
Fig. 27·8

$$p = \frac{e^2 \mathbf{a}^2}{16\pi^2 r^2 c^3} \sin^2 \theta_1.$$

If m be the mass of the particle,

$$m\mathbf{a} = e\mathbf{\mathcal{E}},[1]$$

---

[1] This formula assumes that the electron is free, and, of course, that it is moving with a small velocity. If the electron has a natural frequency, ω_0,

$$m\mathbf{a} = -\frac{e\omega^2}{(\omega_0^2 - \omega^2)}\mathbf{\mathcal{E}},$$

as reference to § 26·4 will show. This reduces to the simple formula when $\omega^2 \gg \omega_0^2$, as is the case with X-rays and γ radiation, and the results obtained in this section apply in general only to the scattering of such short wave radiation.

In the case of low frequency radiation ($\omega_0^2 \gg \omega^2$) a is proportional to ω^2 and consequently the rate of scattering is proportional to ω^4 or to $1/\lambda^4$ (Rayleigh scattering).

and the formula becomes in consequence

$$p = \frac{e^4 \mathcal{E}^2}{16\pi^2 r^2 m^2 c^3} \sin{}^2\theta_1.$$

The average value of p is therefore :

$$I = \frac{e^4 \overline{\mathcal{E}^2}}{16\pi^2 r^2 m^2 c^3} \sin{}^2\theta_1. \quad . \quad . \quad . \quad \textbf{(27·8)}$$

We shall call this average value of p, which we have represented by the letter I, the **intensity** of the scattered radiation at the point of the spherical surface in question. It represents the average rate (per unit area) at which energy travelling in the direction of the arrow in Fig. 27·8 passes through the surface of a sphere of radius r.[1] Similarly, if the beam were polarized with its electric vibrations along the line of the Y axis we should get a corresponding expression for the energy scattered in the direction of the arrow ; but instead of the angle θ_1 we should have the angle θ_2 appearing in the formula. Unpolarized plane waves may be regarded as a superposition of the two cases we have explained ; so that with such a beam

$$I = \frac{e^4}{16\pi^2 r^2 m^2 c^3} \{\overline{\mathcal{E}_x^2} \sin{}^2\theta_1 + \overline{\mathcal{E}_y^2} \sin{}^2\theta_2\},$$

and $\overline{\mathcal{E}_x^2} = \overline{\mathcal{E}_y^2} = \tfrac{1}{2}\overline{\mathcal{E}^2}$. Therefore

$$I = \frac{e^4 \overline{\mathcal{E}^2}}{32\pi^2 r^2 m^2 c^3}\{\sin{}^2\theta_1 + \sin{}^2\theta_2\}. \quad . \quad . \quad \textbf{(27·81)}$$

Let ϕ be the angle between the directions of p (i.e. of the arrow in Fig. 27·8) and of the Z axis, along which the waves are travelling. Then

$$\cos{}^2\phi + \cos{}^2\theta_1 + \cos{}^2\theta_2 = 1,$$

and consequently

$$\sin{}^2\theta_1 + \sin{}^2\theta_2 = 1 + \cos{}^2\phi.$$

On substituting this in **(27·81)** we get :

$$I = \frac{e^4 \overline{\mathcal{E}^2}}{32\pi^2 r^2 m^2 c^3}(1 + \cos{}^2\phi),$$

or $$I = I_{\pi/2}(1 + \cos{}^2\phi), \quad . \quad . \quad . \quad \textbf{(27·82)}$$

---

[1] This statement is only correct when group velocity and phase velocity are identical. They are identical in the cases of practical importance, when the scattering occurs in air or *in vacuo*.

where $I_{\pi/2}$ means the intensity in directions at right angles to that in which the plane incident wave is travelling. It will be noticed that the scattering is strongest in this direction and in the opposite direction ($\phi = 0$ and $\phi = \pi$) and weakest in the directions for which $\phi = \pi/2$. It has the same value in directions making the angles ϕ and $\pi - \phi$ with that of the incident wave.

The intensity of the primary beam, defined as the quantity of energy passing through the unit area per unit time, is, in the units we are using, $c[\mathcal{E}, \mathbf{H}]$ or $c\overline{\mathcal{E}^2}$. If we represent it by I_0 we have :

$$\frac{I_{\pi/2}}{I_0} = \frac{e^4}{32\pi^2 r^2 m^2 c^4}. \quad . \quad . \quad . \quad . \quad (27\cdot83)$$

Since the motions impressed on the scattering electron are in all directions within the plane $\phi = \pi/2$, the radiation scattered along directions in this plane will be completely plane polarized, while it will be unpolarized (if the primary radiation is unpolarized) in the directions $\phi = 0$ and $\phi = \pi$ and partially polarized in the intermediate directions. Apart from the polarization, the scattered radiation will be an exact replica of the primary radiation ; that is to say, the part of the scattered energy included within the range of frequencies, from ω to $\omega + d\omega$ is the same fraction of the corresponding part of I_0 for all frequency ranges. This is obvious from (27·83) which does not contain ω.

If we ascribe the scattering of short-wave radiation by a material particle to the electrons it contains, and if we may regard these electrons as isolated (so that Larmor's formula is valid), and if, finally, we may ignore any consequences of the interference (cf. § 30) of the radiation scattered by different electrons ; then all that has been said about the scattering by a single electron will apply to the scattering by the particle—except of course formula (27·83), the right-hand side of which would have to be multiplied by the number of scattering electrons. Owing partly to the fact that these suppositions are not quite true, and partly to the fact that the classical electromagnetic theory characteristically fails in greater or smaller degree in all phenomena where short wave-lengths are involved, the observed phenomena of the scattering of X-radiation, though resembling very closely what has just been deduced, deviate from it quite appreciably. The observed scattering has not the symmetry demanded by (27·82) ; being more intense in forward directions ($\phi < \pi/2$). Moreover, the scattered radiation is not the exact replica of the primary radiation which the foregoing theory leads us to expect it to be. Scattering brings about an increase in

wave-length **(Compton effect)** which the classical theory is wholly unable to account for. We shall return to these discrepancies in a later volume.

§ 27·9. Coefficient of Scattering of Short Wave-length Radiation

We shall continue with the hypothesis that the scattering of X-rays by materials is due to the electrons they contain, and that these electrons are sufficiently widely separated one from another to justify us in regarding them as isolated. Consequently Larmor's formula and other formulae deduced from it may be made use of. Let us suppose a parallel beam of radiation of very short wave-length passing through some scattering material. Imagine a surface element, dA, perpendicular to the beam (Fig. 27·9) and let us derive an

Fig. 27·9

expression for the energy scattered from the primary beam within the elementary volume $dA\ dx$. We have at once
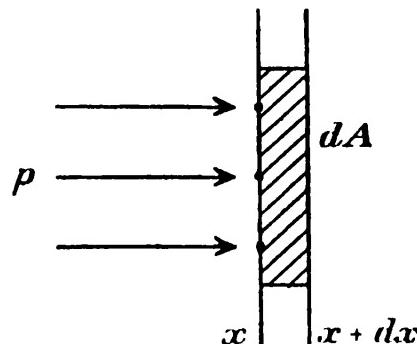
$$\frac{ne^2\mathbf{a}^2}{6\pi c^3}dA\ dx, \quad . \quad . \quad . \quad . \quad . \quad (27\cdot9)$$

in which n is the number of electrons per unit volume. In the case of very short wave-length (high frequency) radiation we may replace $\mathbf{a}$ by $e\mathcal{E}/m$, and if the scattering material is some element we may replace n by

$$n = ZN,$$

where N is the number of atoms of the element in the unit volume and Z the number of electrons in one atom. Hence (27·9) becomes :

$$\frac{NZe^4\mathcal{E}^2}{6\pi m^2 c^3}dA\ dx.$$

Now the intensity, I_0, of the primary beam, i.e. the energy falling on the unit area per unit time, is in our present units $c\mathcal{E}^2$. Therefore the scattered energy is

$$\frac{NZe^4}{6\pi m^2 c^4}\{I_0 dA\}dx; \quad . \quad . \quad . \quad (27\cdot91)$$

which may be written

$$kU dx, \quad . \quad . \quad . \quad . \quad . \quad (27\cdot911)$$

where $U(= I_0 dA)$ is the amount of energy reaching the element per unit time, and k is a constant called the **coefficient of scattering.** We have then

$$k = \frac{NZe^4}{6\pi m^2 c^4}. \qquad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \textbf{(27·92)}$$

The quantity

$$\chi = k/\rho, \qquad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \textbf{(27·921)}$$

where ρ is the density of the material, is sometimes termed the **mass coefficient of scattering** to distinguish it from k which may be called the **volume coefficient of scattering.** Hence

$$\chi = \frac{NZe^4}{6\pi m^2 c^4 \rho}. \qquad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \textbf{(27·93)}$$

Now

$$\rho = NM, \qquad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \textbf{(27·94)}$$

if M means the mass of a single atom, and

$$W = LM, \qquad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \textbf{(27·95)}$$

where W means the atomic weight and L is Loschmidt's number (cf. § **12·6**). Therefore

$$N = \rho L/W. \qquad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \textbf{(27·951)}$$

Substituting this expression for N in **(27·93)** we get for χ :

$$\chi = \left(\frac{e^4 L}{6\pi m^2 c^4} \right) \times \left(\frac{Z}{W} \right). \qquad \cdot \quad \cdot \quad \cdot \quad \textbf{(27·96)}$$

Now e, the charge of an electron, is well known. In the units we are using it is equal to $\sqrt{4\pi} \mid \times 4\cdot774 \times 10^{-10}$, while

$$m = 0\cdot89 \times 10^{-27} \text{ gram,}$$
$$L = 6\cdot02 \times 10^{23}$$

and

$$c = 3 \times 10^{10} \text{ cm./sec.}$$

When these numerical values are substituted in **(27·96)** we find approximately

$$\chi = 0\cdot4 \times \frac{Z}{W}. \qquad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \textbf{(27·97)}$$

The fraction Z/W is the ratio of the number of electrons in the atom to its atomic weight. Barkla found χ experimentally to be very near $0\cdot2$ for several elements of low atomic weight, and thus inferred (correctly as we now know) that in the atoms of these elements the number of electrons is near one-half the atomic weight. On the other hand, the **atomic numbers** of

these elements, i.e. the numbers representing their order in the periodic table, happen also to be close to one-half of the respective atomic weights.   Thus Barkla's experiments on the scattering of X-rays gave the earliest suggestion that the atomic number is identical with the number of electrons in the neutral atom.[1]

§ 28.   Electromagnetic Mass and Momentum—Lorentz-FitzGerald Contraction Hypothesis

We shall now investigate the field due to a charge which is uniformly spread over a spherical surface of radius, R, and which is in motion with a constant velocity, **v**.   Instead of using the Kirchhoff solution of **(27·49)** we shall proceed in another way.   We shall suppose the spherical body to be moving along
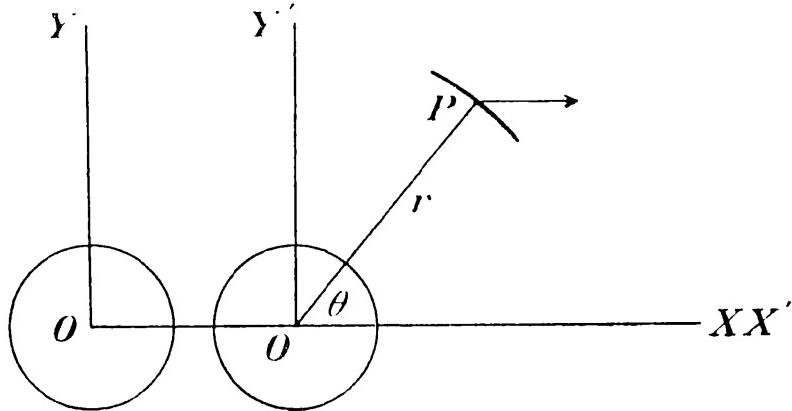
Fig. 28

the X axis of rectangular co-ordinates, and its centre to be at the origin at the instant $t = 0$;   so that at any instant, t, the co-ordinates of its centre are :

$$x = vt, \quad y = 0, \quad z = 0.$$

We shall also use an auxiliary set of rectangular co-ordinates X', Y', Z', with its origin, O', always at the centre of the sphere ; X' coinciding with X and having the same direction ; Y' and Z' being parallel to Y and Z and directed like them.   The equations of transformation from one set of co-ordinates to the other may be written :

$$\left. \begin{aligned} x' &= x - vt, \\ y' &= y, \\ z' &= z, \\ t' &= t. \end{aligned} \right\} \quad . \quad . \quad . \quad . \quad . \quad \textbf{(28)}$$

---

[1] It ought to be pointed out that this refers to *extra-nuclear* electrons. The nucleus may have electrons packed in it ;   in fact we are forced to assume that it has.   The massive nucleus may however be treated as a single entity and its contribution to the scattering is negligible.

Therefore

$$\frac{\partial V}{\partial x} = \frac{\partial V}{\partial x'}\frac{\partial x'}{\partial x} + \frac{\partial V}{\partial y'}\frac{\partial y'}{\partial x} + \frac{\partial V}{\partial z'}\frac{\partial z'}{\partial x} + \frac{\partial V}{\partial t'}\frac{\partial t'}{\partial x}.$$

Consequently by (28)

$$\frac{\partial V}{\partial x} = \frac{\partial V}{\partial x'}$$

and obviously

$$\frac{\partial^2 V}{\partial x^2} = \frac{\partial^2 V}{\partial x'^2} ;$$

so that the part of the equation represented by the laplacian of V remains unmodified in form. To transform $\partial^2 V/c^2 \partial t^2$ we have :

$$\frac{\partial V}{\partial t} = \frac{\partial V}{\partial x'}\frac{\partial x'}{\partial t} + \frac{\partial V}{\partial y'}\frac{\partial y'}{\partial t} + \frac{\partial V}{\partial z'}\frac{\partial z'}{\partial t} + \frac{\partial V}{\partial t'}\frac{\partial t'}{\partial t}.$$

Therefore

$$\frac{\partial V}{\partial t} = - v\frac{\partial V}{\partial x'} + \frac{\partial V}{\partial t'},$$

Now the state of affairs as viewed from the auxiliary co-ordinate system is stationary : i.e. $\partial V/\partial t' = 0$. Therefore

$$\frac{\partial V}{\partial t} = - v\frac{\partial V}{\partial x'},$$

and consequently

$$\frac{\partial^2 V}{\partial t^2} = v^2\frac{\partial^2 V}{\partial x'^2}.$$

On substituting in (27·49) we get :

$$\left(1 - \frac{v^2}{c^2}\right)\frac{\partial^2 V}{dx'^2} + \frac{\partial^2 V}{\partial y'^2} + \frac{\partial^2 V}{\partial z'^2} = - \rho, \quad . \quad . \quad \textbf{(28·01)}$$

an equation which resembles Poisson's equation (18·721). We shall continue to use the notation of § 27·2 and write

$$(1 - v^2/c^2) = 1/\gamma^2.$$

Consequently (28·01) becomes :

$$\frac{\partial^2 V}{\partial \xi^2} + \frac{\partial^2 V}{\partial \eta^2} + \frac{\partial^2 V}{\partial \zeta^2} = - \rho, \quad . \quad . \quad \textbf{(28·02)}$$

if

$$\xi = \gamma x',$$
$$\eta = y',$$
$$\zeta = z'.$$

This equation has exactly the form of Poisson's equation; but it has not quite the same meaning on account of the fact that the distances, ξ, are not distances measured along the X' axis in the ordinary way. The simplicity of Poisson's equation and our acquaintance with solutions of it, make it worth while to investigate the consequences of identifying ξ with an X co-ordinate. We shall in fact introduce the hypothesis that the naïve assumption

$$x' = x - vt$$

in (28) is faulty, and that measurements along the X' axis are correctly represented by

$$\xi = \gamma(x - vt).$$

If therefore $x_2 - x_1$ represent some length measured along the X axis of the original system of co-ordinates, then the same length measured in the system $X'Y'Z'$ is

$$\xi_2 - \xi_1 = \gamma(x_2 - x_1).$$

The new hypothesis therefore amounts to the assumption that **measurements of length made in the direction of the X' or X axis will be greater when the measuring scale is moving with the body measured, i.e. is at rest relatively to the body, than such measurements made with a scale relatively to which the body is in motion in this direction.** This is the famous **contraction hypothesis** introduced independently by FitzGerald and H. A. Lorentz to account (as it did) for the negative result of the experiments of Michelson, and Michelson and Morley, on the velocity of light (cf. §§ 31·3 and 33·4). Doubtless too the hypothesis was suggested to Lorentz by the fact that (28·02) has the same mathematical form as Poisson's equation. Before it becomes quite identical with Poisson's equation it requires a further, mathematically trivial, modification. **The quantity, ρ, is the electric density as understood without the FitzGerald-Lorentz hypothesis.** Since the volume

$$d\xi\ d\eta\ d\zeta = \gamma\ dx\ dy\ dz,$$

the charge in the element $dx\ dy\ dz$ is

$$\rho\ dx\ dy\ dz = \frac{\rho}{\gamma}\ d\xi\ d\eta\ d\zeta,$$

and the density, ρ', referred to the $\xi,\ \eta,\ \zeta$ co-ordinates is consequently

$$\rho' = \rho/\gamma. \qquad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \textbf{(28·03)}$$

We have therefore for the final form of the equation :

$$\frac{\partial^2(V/\gamma)}{\partial\xi^2} + \frac{\partial^2(V/\gamma)}{\partial\eta^2} + \frac{\partial^2(V/\gamma)}{\partial\zeta^2} = -\rho' \qquad . \quad \textbf{(28·04)}$$

which is identifiable in all respects with the electrostatic equation of Poisson.

The adoption of the FitzGerald-Lorentz hypothesis has introduced some uncertainty about the shape of the surface over which the electric charge is spread. We shall make this definite by describing it as spherical when measured by a scale relatively to which it is at rest. It is a spherical surface in the (ξ, η, ζ) co-ordinates, and we shall suppose that it is this surface over which the electricity is uniformly distributed (surface density constant).

The appropriate solution of **(28·04)** is

$$V/\gamma = \frac{e}{4\pi r}, \quad . \quad . \quad . \quad . \quad . \quad . \quad \textbf{(28·05)}$$

where

$$r = (\xi^2 + \eta^2 + \zeta^2)^{1/2}$$

is the distance from the centre of the sphere (referred to the (ξ, η, ζ) co-ordinates) of the point where V is the required potential. It must not be forgotten that V is the scalar potential referred to the original X, Y, Z co-ordinates ; although in **(28·05)** it is expressed in terms of (ξ, η, ζ).

We have therefore

$$\left. \begin{array}{l} V = \dfrac{\gamma e}{4\pi r}, \\[3mm] \mathbf{A} = A_x = \dfrac{\gamma e \mathrm{v}}{4\pi c r}. \end{array} \right\} \quad . \quad . \quad . \quad \textbf{(28·06)}$$

We shall now calculate the electromagnetic momentum per unit volume at some point P. For this purpose we need to evaluate the X component of Poynting's vector. This is

$$c\{\mathcal{E}_y H_z - \mathcal{E}_z H_y\}$$

or

$$c\mathcal{E}_y H_z,$$

if, as we shall suppose, the point is in the plane XY, since only the Z component of $\mathbf{H}$ differs from zero. The momentum per unit volume is therefore at such a point

$$\frac{1}{c}\mathcal{E}_y H_z. \quad . \quad . \quad . \quad . \quad \textbf{(28·061)}$$

Now
$$\mathcal{E}_y = -\frac{\partial V}{\partial y} - \frac{1}{c}\frac{\partial A_y}{\partial t},$$

or
$$\mathcal{E}_y = -\frac{\partial V}{\partial y}.$$

Now it will be observed that $\partial/\partial y \equiv \partial/\partial y' \equiv \partial/\partial \eta$. Therefore

$$\mathcal{E}_y = \frac{\gamma e}{4\pi r^2}\frac{\partial r}{\partial \eta} = \frac{\gamma e}{4\pi r^2}\sin\theta. \quad . \quad . \quad . \quad (28\cdot07)$$

The $\mathcal{E}_y$ on the left of the equation means of course the value appropriate to the X, Y, Z system; but the terms in which it is expressed, e.g. r and $\sin\theta$, have the values appropriate to the (ξ, η, ζ) system. Similarly,

$$H_z = \frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y},$$

or
$$H_z = -\frac{\partial \mathbf{A}}{\partial y} = -\frac{\partial \mathbf{A}}{\partial y'} = -\frac{\partial \mathbf{A}}{\partial \eta}.$$

Therefore

$$H_z = \frac{\gamma e \mathbf{v}}{4\pi c r^2}\sin\theta. \quad . \quad . \quad . \quad . \quad . \quad (28\cdot071)$$

Combining this with $(28\cdot061)$ and $(28\cdot07)$ we get for the X component of the momentum per unit volume:

$$\frac{\gamma^2 e^2 \mathbf{v}}{16\pi^2 c^2 r^4}\sin^2\theta. \quad . \quad . \quad . \quad . \quad (28\cdot08)$$

This is the momentum per unit volume at a point, P, in the XY plane. The type of symmetry of the problem shows us that it is valid for all points when θ means the angle (in the ξ, η, ζ system) between the line r and the ξ axis. A volume element in the (ξ, η, ζ) system may be put in the form:

$$r^2 \sin\theta \, d\theta \, d\phi.$$

It will correspond to

$$\frac{r^2 \sin\theta \, d\theta \, d\phi}{\gamma} \quad . \quad . \quad . \quad . \quad (28\cdot081)$$

in the XYZ system. If we multiply $(28\cdot08)$ by $(28\cdot081)$ and integrate between the limits:

$$R \text{ and } \infty \text{ for } r,$$
$$0 \text{ and } \pi \text{ for } \theta$$

and
$$0 \text{ and } 2\pi \text{ for } \phi;$$

we get
$$M = \frac{\gamma e^2 \mathbf{v}}{16\pi^2 c^2} \int \int \int \frac{dr \, \sin^3\theta \, d\theta \, d\phi}{r^2},$$

or

$$M = \gamma \cdot \frac{e^2}{6\pi c^2 R} \cdot \mathbf{v}, \qquad . \qquad . \qquad . \qquad . \qquad . \qquad . \quad (28 \cdot 09)$$

where R is the radius of the sphere.

If we divide this formula by $\mathbf{v}$, the velocity of the sphere, we obtain an expression, namely

$$\gamma \frac{e^2}{6\pi c^2 R}, \qquad . \qquad . \qquad . \qquad . \qquad . \qquad . \quad (28 \cdot 091)$$

for the part of the mass which is due to the possession by the body of a charge. The factor, γ, approaches the limit 1 for small velocities and **28·091** then becomes identical with **26·14**, when we take into account the fact that it is expressed in terms of the Lorentz-Heaviside unit of charge. We find again, as we might have anticipated, the relationship

$$m = \gamma m_0,$$

where m is the mass for the velocity $\mathbf{v}$, and m_0 the mass for the limit $\mathbf{v} = 0$.

BIBLIOGRAPHY

H. A. WILSON : Modern Physics. (Blackie.)

O. W. RICHARDSON : Electron Theory of Matter. (Cambridge.)

H. A. LORENTZ : Versuch einer Theorie der elektrischen und optischen Erscheinungen in Bewegten Körpern. (Teubner, Leipzig.) Lectures on Theoretical Physics, Vol. III. (Macmillan.)

LEIGH PAGE : An Introduction to Electrodynamics. (Ginn & Co.)

CHAPTER XI

GEOMETRICAL OPTICS

§ 28·1. HISTORICAL NOTE

SPECIFICALLY optical phenomena are part of the larger group of electromagnetic phenomena. This was the great discovery of Clerk Maxwell.[1] All that has been established in the preceding chapters about electromagnetic waves, their reflexion and refraction, polarization and dispersion, applies to light waves, which are in fact electromagnetic waves ; and the whole theory of electromagnetism is relevant to optics. The latter owes its *quasi* independence and its separate terminology to visual experimental methods and similar methods developed out of them. The efforts of Maxwell himself and his predecessors were directed, in the spirit of an earlier scientific ideal, towards a consistent mechanical theory of light. Two types of such optical theories have been prominent in the history of physics before the time of Maxwell : corpuscular and undulatory. The former, which is very ancient and appears to have been held by Pythagoras in the sixth century B.C., represented light as a stream of minute corpuscles shot out from the luminous body. Newton carried out an extensive series of optical experiments, from which he developed a theory of the corpuscular kind. He does not appear to have enunciated the specific hypothesis that light consists of corpuscles. Indeed, his work on *Opticks* (1704) begins with the characteristic disavowal : *My Design in this Book is not to explain the Properties of Light by Hypotheses, but to propose and prove them by Reason and Experiment.* . . . He uses the term **ray** (cf. § 28·2) in a sense resembling that of **corpuscle :** *The least Light or part of Light, which may be stopt alone, or do or suffer anything alone, which the rest of the Light doth not or suffers not, I call a Ray of Light.* On pages 57 and 58 of the same treatise he treats the rays as if they really were material particles, subject, in the neighbourhood of the boundary

[1] Clerk Maxwell, ' A Dynamical Theory of the Electromagnetic Field ', *Phil. Trans.*, Vol. 155, 1864.

between two media, to forces normal to the boundary, and applies a principle which is identical with the energy principle. He therefore does tacitly, if inadvertently, introduce hypotheses into his theory.

A remarkable feature of Newton's optical theory, and one which was prophetic of quite recent theoretical views on the nature of light, was the assumed existence in each ray (corpuscle) of a periodic variation, or change of state. This Newton inferred from his observations of interference phenomena (Newton's rings).

An undulatory theory of light was developed by Christian Huygens (1629–1695), a contemporary of Newton, and published in his *Traité de la Lumière* (1690). He regarded light waves as longitudinal, and was consequently unable to find an explanation of the phenomenon of polarization. The most interesting and enduring feature of his theory is the principle, since known as **Huygens' principle,** which he successfully applied to account for the elementary laws of reflexion and refraction, not only in isotropic media, but also in uniaxial crystals. Newton's theory, however, dominated optics till the close of the eighteenth century, when the undulatory theory was revived by Thomas Young (1773–1829). Its further development was chiefly the work of Augustin Jean Fresnel (1788–1827), who adopted Young's suggestion that light waves were of the transverse type, and propagated by a medium (aether) with properties resembling those of an elastic solid. He supposed this luminiferous medium to penetrate material media, and to have everywhere the same elasticity (rigidity) ; so that the different velocities of propagation of light waves in different media had to be ascribed to the differences in the aether density or concentration in them. The directions of the displacement of the particles or elements of Fresnel's luminiferous medium, when traversed by a plane polarized wave, are perpendicular to the conventional plane of polarization (and therefore to the plane of incidence in the case of light polarized by reflexion at a glass surface). The displacement therefore has a position in his theory like that of the electric displacement in Maxwell's theory. Franz Neumann (1798–1895) and James MacCullagh (1809–1847) developed independently of one another undulatory theories in which the displacements of the medium traversed by a plane polarized wave were *in the conventional plane of polarization.* Their efforts and those of their contemporaries were directed towards establishing the undulatory theory of light on a sound dynamical basis, and the most interesting feature of MacCullagh's theory, apart from its high degree of logical coherence, lies in the fact that its mathematical form resembles that of Maxwell's theory much more closely than it

does that of the propagation of waves in an ordinary elastic solid.

Fresnel's great achievements were the development of the sine and tangent formulae of §§ 25·5 and 25·6 and the consequent explanation of Brewster's law (25·7) ; a very complete treatment of diffraction phenomena by the use of an improved form of Huygens' principle, and a complete theory of the phenomena of polarization and propagation in crystalline media. Most of the consequences of his optical theory are in excellent accord with observational results, and in some instances he correctly predicted the results of experiments which were not carried out till long after his death—as, for example, the experiments of Fizeau on the influence of the motion of a material medium on the measured velocity of light in it, and the negative result of the experiment of Airy carried out to ascertain whether the aberration of light as observed by a telescope filled with some transparent material medium (water) differed from that observed with an ordinary telescope, §§ 33·2 and 33·3. Although the dynamical foundations of all his work are characteristically obscure, it is by far the most important contribution to optical science before the advent of the electromagnetic theory of Clerk Maxwell.

§ 28·2. Geometrical and Physical Optics

Starting out with the view that light is an undulatory phenomenon, but without making any use for the moment of the features peculiar to the electromagnetic theory, we note that a plane harmonic wave may be described by the formula :

$$s = A \cos 2\pi\left(\frac{t}{\tau} - \frac{z}{\lambda}\right), \qquad . \quad . \quad . \quad . \quad (28\cdot2)$$

which needs no elucidation, and bear in mind the explanation at the beginning of § 27·1. We therefore regard the wave (28·2) as the limit approached by the group consisting of a superposition of plane waves of wave-lengths included between λ and $\lambda + d\lambda$ with periods between corresponding limits τ and $\tau + d\tau$, and travelling in directions included within some narrow solid angle, $d\Omega$, containing the Z axis, when $d\lambda$ and $d\tau$ become negligible compared with λ and τ respectively. It is there explained that this limiting case implies that λ is negligible compared with the linear dimensions of the group, and that similarly τ is negligible compared with the time it takes to travel through its own length. If therefore we imagine a limited

portion to be cut out, as it were, from such a beam of light, it will continue to travel as if it still formed part of it, its shape and phase velocity persisting, as if it were, intact in the original beam, *provided* of course that its dimensions are large compared with the wave-length. This condition may be satisfied while the dimensions of this limited portion of the beam are so small compared with those of the region we are investigating, that we may regard it as practically a point. It will then in fact be one of Newton's rays—with, however, the important difference that in general *two velocities* are associated with it : one being its own proper velocity with which its configuration advances (group velocity), and the other being the velocity with which the crests and troughs advance (phase velocity). The smallness of its dimensions enables us to regard its path as a line. It is this line which we, following the established usage, shall term a **ray of light,** and not the group itself as Newton did.

The part of optics dealing with the phenomena falling within the limiting conditions we have described is called **geometrical optics**—a term which describes the character of the mathematical methods appropriate for this part of the subject. The part of optics lying outside this region is called **physical optics.** The former phenomena are of course merely limiting cases of the latter, and there is no sharp boundary separating them.

§ 28·3. OPTICAL PATH

While continuing to deal with light of a definite period, i.e. periods within the range $\tau - \tau + d\tau$, we shall follow its course through different media with each of which a wave-length is associated which is determined by τ and the nature of the medium. We shall use the term **optical path** for the course of a Newtonian ray (corpuscle or group) from a point A to another B, or for any path from A to B which may be contemplated as a possible course for the light to take, and shall base the definition of its length, or the **optical distance** between the two points, on the principle that equal paths contain the same number of waves, and are therefore traversed in equal times—*not by the light*—but by the phase (crests or troughs). Consider, for example, the path from A to B in Fig. 28·3, extending through three different media, indicated by the numbers 1, 2 and 3, and crossing the boundaries between
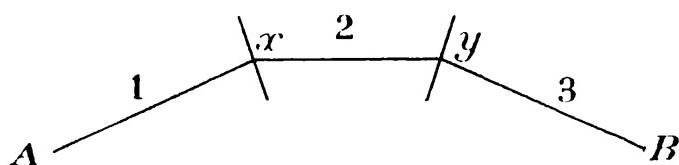
Fig. 28·3

them at x and y. We may employ as a measure of the optical path

$$l_1/\lambda_1 + l_2/\lambda_2 + l_3/\lambda_3. \quad . \quad . \quad . \quad (28·3)$$

where l_1 is the distance (in the ordinary sense of the term) from A to x, l_2 that from x to y and l_3 the remaining distance ; while λ_1, λ_2 and λ_3 are the corresponding wave-lengths.

The refractive indices, n_1, n_2 and n_3, of these media mean c/u_1, c/u_2 and c/u_3 respectively, where c is the velocity of waves of the same period *in vacuo* and u_1, u_2 and u_3 the respective phase velocities in 1, 2 and 3. Now

$$c = \lambda/\tau, \quad u_1 = \lambda_1/\tau, \quad u_2 = \lambda_2/\tau \text{ and } u_3 = \lambda_3/\tau,$$

λ being the wave-length of such waves *in vacuo* ; therefore

$$\lambda = n_1\lambda_1 = n_2\lambda_2 = n_3\lambda_3,$$

and so the expression **(28·3)** may be written :

$$\left. \begin{array}{c} \{n_1l_1 + n_2l_2 + n_3l_3\}/\lambda, \\[4pt] \dfrac{1}{\lambda}\displaystyle\int_A^B ndl. \end{array} \right\} \quad . \quad . \quad . \quad \textbf{(28·31)}$$

or more generally

This is the optical distance between A and B measured by counting the number of waves in it. It is most usual to take as its measure the distance *in vacuo* containing the same number of waves ; that is to say, the distance

$$n_1l_1 + n_2l_2 + n_3l_3,$$

or, as we may more generally express it,

$$\int_A^B ndl. \quad . \quad . \quad . \quad . \quad . \quad \textbf{(28·32)}$$

§ 28·4. Fermat's Principle

Pierre de Fermat (1601–1665), a famous French mathematician, enunciated the important principle (in geometrical optics) that the path selected by the light in travelling from one point to another is that which can be traversed in the least time. The discussion in **§ 28·2** shows that this simple statement is ambiguous. The natural interpretation of ' least time ' would be ' least time required by a group, corpuscle or Newtonian ray ',

and with this interpretation the expression for the time would be

$$\int_A^B \frac{dl}{v}, \qquad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \textbf{(28·4)}$$

where v is the *group* velocity. The other interpretation makes the time

$$\int_A^B \frac{dl}{u}, \qquad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \textbf{(28·41)}$$

where u is the associated *phase* velocity. The corresponding statements of the principle would therefore be

$$\delta \int_A^B \frac{dl}{v} = 0 \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \textbf{(28·42)}$$

with the former interpretation and

$$\delta \int_A^B \frac{dl}{u} = 0 \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \textbf{(28·43)}$$

with the latter, and of these two it is the latter which correctly expresses Fermat's principle. If we interpret 'least time' literally the principle is not in general true. The path of a ray may be one for which the time is greatest ; or there may be an infinite number of paths for all of which the time is the same ; or again the path may be one for which the time is neither a maximum nor a minimum, as, for example, when a ray from a fixed point, A, to another, B, is reflected at a saddle-shaped surface. As expressed by equation **(28·43)** the principle is, however, true, and it is better to speak of it as the **principle of stationary time.** The integral **(28·41)**, which represents the (phase) time along some path from A to B, is an analytic function of certain independent variables, which we may call ϕ, ψ, etc. Any increment of this time due to a change in these variables from ϕ_0 to $\phi_0 + \Delta\phi$, ψ_0 to $\psi_0 + \Delta\psi$, etc., has the general form :

$$R\Delta\phi + S\Delta\psi + \ldots + \begin{Bmatrix} \text{terms involving higher powers} \\ \text{and products of } \Delta\phi, \Delta\psi, \text{ etc.,} \\ \text{two or more at a time} \end{Bmatrix},$$

the coefficients R and S being functions of ϕ_0, ψ_0, etc., and independent of $\Delta\phi$, $\Delta\psi$, etc. Now the symbol δ simply refers

to the portion $R\Delta\phi + S\Delta\psi + \ldots$ of the increment, in the limiting case where $\Delta\phi$, $\Delta\psi$, etc., are evanescent. So that

$$\delta\int_A^B \frac{dl}{u} = R\delta\phi + S\delta\psi + \ldots$$

Since ϕ, ψ, etc., are independent variables, $\delta\phi$, $\delta\psi$ . . . are arbitrary increments, and therefore the statement

$$\delta\int_A^B \frac{dl}{u} = 0$$

implies that R, S, etc., are individually zero.

Now we are dealing with light of some given period, τ, by hypothesis, and τ is the time required by the wave crest, or some given phase, to travel the length of a wave. The time required therefore for the path between two given points, A and B, is equal to the product of τ and the number of waves included in it ; in fact it is equal to the product of τ and the length of the optical path as measured by the number of waves it contains, and may be generally expressed by

$$\frac{\tau}{\lambda}\int_A^B n\,dl, \quad . \quad . \quad . \quad . \quad . \quad . \quad \text{(28·44)}$$

where λ is the wave-length *in vacuo*. Fermat's principle may therefore be stated in the following way :

$$\delta\frac{\tau}{\lambda}\int_A^B n\,dl = 0,$$

or, since τ/λ is not subject to variation,

$$\delta\int_A^B n\,dl = 0, \quad . \quad . \quad . \quad . \quad . \quad \text{(28·45)}$$

where A and B are the terminal points of the path. The principle therefore affirms that the length of the optical path traversed by the light in passing from A to B has a stationary value.

Its justification emerges from the fact that, since the group and phase velocities coincide in direction, the track of a group—this is only another name for a ray of light—from a point, A, to another point, B, is everywhere normal to the surfaces of

constant phase (wave surfaces) which it intersects. Let AB in Fig. 28·4 represent the actual track or path from A to B. Its optical length is expressible in terms of the parameters, ϕ, ψ, . . . Shall we say it is expressed by $\phi = \phi_0$, $\psi = \psi_0$, etc. In the figure are shown surfaces of constant phase, aa', bb', cc', etc., which the ray AB cuts orthogonally at a, b, c, . . . and also another path a', b', c', . . . cutting the same surfaces orthogonally. This latter path also represents the course of a ray of light, though not necessarily one from A to B. In consequence of the definition of optical distance, we have :

FIG. 28·4

$(ab) = (a'b')$; $(bc) = (b'c')$; $(cd) = (c'd')$; . . . exactly, when (ab), (bc), $(a'b')$, etc., mean optical distances, not distances in the ordinary sense of the term. Now the path from A to B (represented by the broken line), for which the parameters ϕ, ψ, . . . have the respective values, $\phi_0 + \delta\phi$, $\psi_0 + \delta\psi$, . . . must cut the wave surfaces at angles which deviate from orthogonality by small amounts only : at most of the order of $\delta\phi$, $\delta\psi$, . . . and in consequence the optical distance along it, between the surfaces bb' and cc', for example, will differ from the optical distance, (bc), by a second order small quantity at most. It follows therefore that the whole length of this neighbouring optical path can only differ from that of the actual ray from A to B by a second or higher order small quantity. Hence

$$\delta \int_A^B ndl = R'\delta\phi + S'\delta\psi + \ldots = 0,$$

with an arbitrary choice of $\delta\phi$, $\delta\psi$, . . .

§ 28·5. Laws of Reflexion and Refraction

The elementary laws of refraction and reflexion may be regarded as immediate consequences of Fermat's principle. We shall consider in some detail the path from a point A in the medium 1 to another point B in the medium 2 (Fig. 28·5). This path must necessarily consist of two straight lines, which may or may not be parts of one and the same straight line. This is an obvious consequence of Fermat's principle. Let us refer the
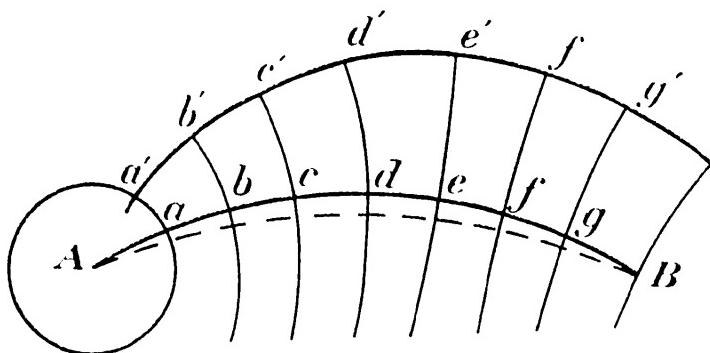
path to rectangular axes of co-ordinates. We may regard the part of the surface of separation between the two media, in the neighbourhood of the point where the path crosses it, as plane, and represent it by

$$ax + by + cz + d = 0. \quad . \quad . \quad . \quad (28\cdot5)$$

The co-ordinates of the point where the light crosses the boundary (point of incidence) we shall represent by (x, y, z); those of A and B by (x_1, y_1, z_1) and (x_2, y_2, z_2) respectively. Let r_1 and r_2 represent the *actual* (not the optical) lengths of the parts of the path in 1 and 2 respectively, and n_1 and n_2 the respective refractive indices.

If L be the length of the optical path,

$$L = n_1 r_1 + n_2 r_2,$$

or

Fig. 28·5

$$L = n_1 \{ (x - x_1)^2 + (y - y_1)^2 + (z - z_1)^2 \}^{1/2}$$
$$+ n_2 \{ (x - x_2)^2 + (y - y_2)^2 + (z - z_2)^2 \}^{1/2},$$

and consequently

$$\delta L = \left\{ \frac{n_1(x - x_1)}{r_1} + \frac{n_2(x - x_2)}{r_2} \right\} \delta x$$

$$+ \left\{ \frac{n_1(y - y_1)}{r_1} + \frac{n_2(y - y_2)}{r_2} \right\} \delta y$$

$$+ \left\{ \frac{n_1(z - z_1)}{r_1} + \frac{n_2(z - z_2)}{r_2} \right\} \delta z = 0. \quad . \quad (28\cdot51)$$

This is required by Fermat's principle. The small variations δx, δy, δz are not independent, since they are necessarily in the surface (28·5), and are therefore subject to the condition :

$$a\delta x + b\delta y + c\delta z = 0. \quad . \quad . \quad (28\cdot511)$$

If we eliminate δx we get a single equation of the form :

$$\alpha\delta y + \beta\delta z = 0,$$
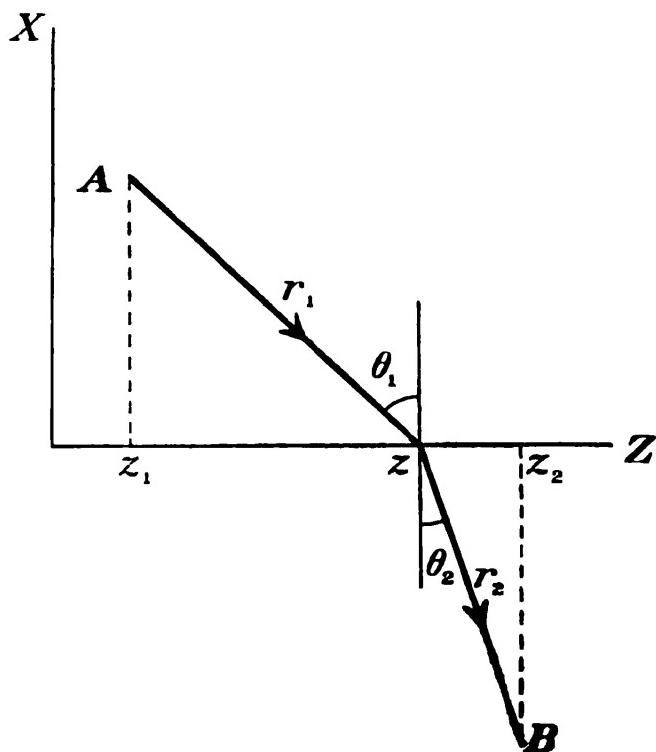
in which δy and δz are arbitrary, and therefore we conclude that

$$\left.\begin{array}{c} \alpha = 0, \\ \beta = 0. \end{array}\right\} \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \textbf{(28·52)}$$

These equations describe the path of the light.

We can carry out this process most conveniently, and get the final statement **(28·52)** in its simplest form, if we place our axes of co-ordinates so that A and B are in the plane of X and Z, and therefore y_1 and y_2 are both zero. We may also, while maintaining this condition, have them so placed that the refracting surface is in the plane YZ, with the consequence that **(28·511)** becomes

$$\delta x = 0.$$

Equations **(28·52)** thus become

$$\left.\begin{array}{c} \left(\dfrac{n_1}{r_1} + \dfrac{n_2}{r_2}\right)y = 0, \\[2mm] \dfrac{n_1(z - z_1)}{r_1} + \dfrac{n_2(z - z_2)}{r_2} = 0. \end{array}\right\} \quad \cdot \quad \cdot \quad \cdot \quad \textbf{(28·53)}$$

Now $n_1/r_1 + n_2/r_2$ is necessarily different from zero. Therefore $y = 0$, which means that the point of incidence is in the XZ plane, and this is therefore a plane containing the incident ray, r_1, the refracted ray, r_2, and the normal at the point of incidence. Reference to Fig. 28·5 will make it clear that

$$(z - z_1)/r_1 = \sin \theta_1,$$

and

$$(z_2 - z)/r_2 = \sin \theta_2,$$

where θ_1 and θ_2 are respectively the angles of incidence and refraction, according to the usual definitions. Hence the second equation **(28·53)** is equivalent to

$$\frac{\sin \theta_1}{\sin \theta_2} = \frac{n_2}{n_1},$$

which is the remaining law of refraction.

The laws of reflexion may be deduced in a similar way from Fermat's principle.

§ 28·6. Formation of Images by Spherical Refracting Surfaces and Pencils of Paraxial Rays

In Fig. 28·6, PM is part of a spherical refracting surface separating two media, 1 and 2, with refractive indices n_1 and n_2

respectively. The centre of curvature of the surface is at C. Let O be a point source of light (a point object). We shall term the straight line through O and C the **axis** and shall deal with the case where the area of the refracting surface reached by light rays is so restricted that all the rays lie in a very small solid angle ; so small that the sine or tangent of the angle which any ray makes with the axis need not be distinguished from the angle itself.

If now we pick out any other point, I, on the axis, Fermat's principle (or the pair of laws of refraction deduced from it) indicates that there is in general only one possible path from O to I, namely along the axis. None of the rays except the axial one can travel from O to I in general. If, however, the points

Fig. 28·6

O and I be suitably chosen, all the rays travelling from O to the refracting surface will on refraction pass through I, or at any rate through a region round about I so small that we need not distinguish it from that point. We are of course assuming the existence of those limiting conditions which mark the province of geometrical optics. The point I is then called the **real image** of O. We shall use the letters u, v and r to represent respectively the distances of the object, image and centre of curvature from the point, P, where the axis cuts the refracting surface, adopting the convention that distances measured in the opposite direction to that of the axial ray are positive ; so that in the case represented in Fig. 28·6

$$u = (OP),$$
$$v = - (IP),$$
and
$$r = - (CP),$$

if (OP), (IP) and (CP) mean the spacial intervals between the points indicated.

It is easily seen that in order that all the rays from O shall pass through I (at least approximately) the following formula must hold :

$$\frac{n_2/n_1}{v} - \frac{1}{u} = \frac{n_2/n_1 - 1}{r} . \quad . \quad . \quad . \quad \textbf{(28·6)}$$

This formula is necessary, but it does not follow that, when
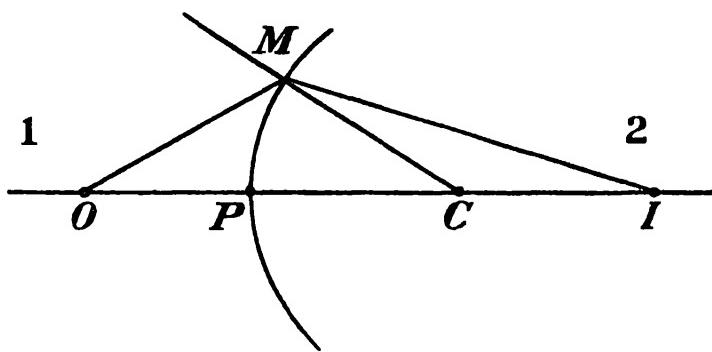
it is satisfied, the point I is one through which the rays of light from O pass ; but if they do not, the directions of the refracted rays in 2, if produced backwards, will pass through a point I. When this case occurs I is called a **virtual image.**

The formula **(28·6)** may be written in the form :

$$\frac{f_2}{v} + \frac{f_1}{u} = 1, \quad . \quad . \quad . \quad . \quad . \quad \textbf{(28·61)}$$

where

and

$$\left. \begin{aligned} f_1 &= -\frac{r}{n_2/n_1 - 1}, \\[2mm] f_2 &= \frac{n_2 r/n_1}{n_2/n_1 - 1}. \end{aligned} \right\} \quad . \quad . \quad . \quad \textbf{(28·611)}$$

The axial points which are distant f_1 and f_2 from P are called the **first** and **second principal foci** respectively, and f_1 and f_2 are called the **first** and **second focal lengths** of the refracting system.

It is obvious that **(28·61)** may be written in the form :

$$xx' = f_1 f_2, \quad . \quad . \quad . \quad . \quad . \quad . \quad \textbf{(28·62)}$$

in which

and

$$\left. \begin{aligned} x &= u - f_1, \\ x' &= v - f_2 ; \end{aligned} \right\} \quad . \quad . \quad . \quad \textbf{(28·621)}$$

so that x means the distance of the object point from the first principal focus, while x' means the distance of the image point from the second principal focus, the same convention being used as in the case of r, u and v.

Now consider an object point, A, near O and in the plane through O perpendicular to the axis (Fig. 28·61). It will have its image, B, near I and approximately in the plane through I perpendicular to the axis. To show that this is the case we have only to construct a new axis through A and the centre of curvature. We then see that the relationship between (AP) and (PB) is the same as that between (OP) and (PI). Consequently the length (AP) differs from (OP) by a small quantity
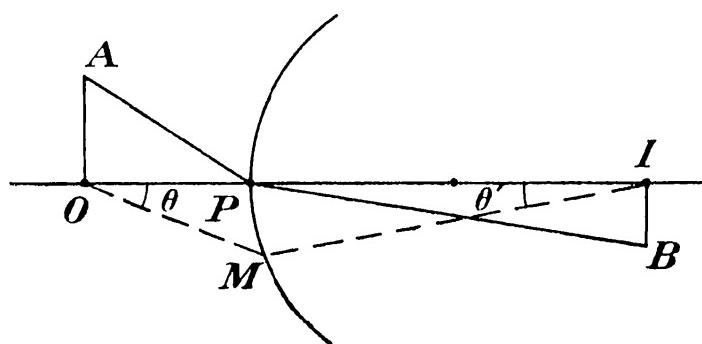
FIG. 28·61

of the second order, since (OA) is a first order small quantity. Similarly, (PB) differs from (PI) by a small quantity of the second order. Let us represent distances in the object space, measured perpendicularly to the axis, by y, using the positive sign for the upward direction ; and similarly y' for distances in the image space. By considering the ray APB we easily see that

$$\frac{y}{y'} = \frac{n_2}{n_1}\frac{u}{v}, \qquad . \quad . \quad . \quad . \quad . \quad (28\text{·}63)$$

and by (28·611) this is equivalent to

$$\frac{y}{y'} = -\frac{f_2}{f_1}\frac{u}{v}. \qquad . \quad . \quad . \quad . \quad . \quad (28\text{·}64)$$

If we consider such a path as OMI, the parts of which in the object and image spaces make the respective angles θ and θ' with the axis, we find that (28·63) is equivalent to

$$n_1 y \theta = n_2 y' \theta'. \quad . \quad . \quad . \quad . \quad . \quad (28\text{·}65)$$

By making use of (28·62) and (28·621), we can express (28·64) in the forms :

$$\left. \begin{array}{l} \dfrac{y}{y'} = -\dfrac{x}{f_1}, \\[2mm] \dfrac{y}{y'} = -\dfrac{f_2}{x'}. \end{array} \right\} \quad . \quad . \quad . \quad . \quad . \quad (28\text{·}66)$$

If instead of a single refracting surface we have two such surfaces a and b (Fig. 28·62) separated by an interval, t, the first one, a, forms an image of O, and the surface b forms an image of this image. For this final image we find, by applying the methods appropriate to a single refracting surface twice over,
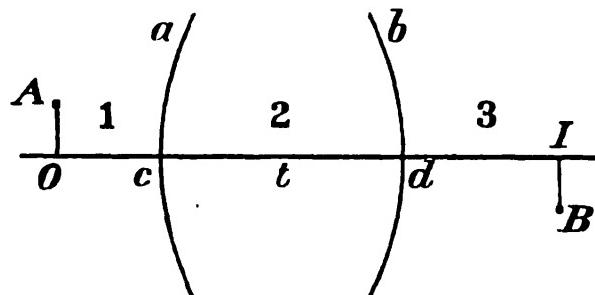
Fig. 28·62

$$(M - \phi)\,(N - \phi') = \text{const.}, \quad . \quad . \quad . \quad (28\text{·}67)$$

where M means the distance Oc of the point O from the surface, a, provided with its proper sign, and N is similarly the distance dI of the final image I from the second refracting surface, b.

The symbols ϕ and ϕ' are abbreviations for

$$\phi = \frac{f^a{}_1 f^b{}_1 - t f^a{}_1}{f^b{}_1 - t - f^a{}_2},$$

$$\phi' = \frac{f^a{}_2 f^b{}_2 + t f^b{}_2}{f^a{}_2 + t - f^b{}_1},$$

$f^a{}_1$, for example, meaning the first focal length of the spherical surface, a, $f^b{}_2$ the second focal length of the surface, b, and so on ; while t means the absolute value of the distance between c and d.

The formula (28·67) indicates that there is a first principal focus of the system of two refracting surfaces at the point $M = \phi$, and a second one at the point, $N = \phi'$. It is therefore appropriate to express it in the form :

$$xx' = \text{const.},$$

or if we make a suitable choice of the points from which we may agree to measure the respective focal lengths of the system,

$$xx' = f_1 f_2. \qquad . \qquad . \qquad . \qquad . \qquad . \qquad \textbf{(28·68)}$$

We may now show that

$$\frac{y}{y'} = \frac{f^b{}_2}{N} \frac{(M f^a{}_2 + M t - t f^a{}_1)}{f^a{}_1 f^b{}_1},$$

from which we get

$$\frac{y}{y'} = - \frac{(M - \phi)}{f^a{}_1 f^b{}_1 / (f^b{}_1 - t - f^a{}_2)},$$

which may be written

$$\frac{y}{y'} = - \frac{x}{f_1},$$

if we agree to fix the point from which f_1 is to be measured in such a way that

$$f_1 = \frac{f^a{}_1 f^b{}_1}{f^b{}_1 - t - f^a{}_2}. \qquad . \qquad . \qquad . \qquad . \qquad \textbf{(28·681)}$$

This fixes the value of f_2 also, since $f_1 f_2$ has to be equal to a certain constant (28·67), and thus it also fixes the point from which f_2 must be measured. We easily find for f_2,

$$f_2 = \frac{f^a{}_2 f^b{}_2}{f^a{}_2 + t - f^b{}_1}. \qquad . \qquad . \qquad . \qquad . \qquad \textbf{(28·682)}$$

The points from which we have measured f_1 and f_2 are called the **first** and **second principal points** respectively, while the planes perpendicular to the axis at these points are the **first** and **second principal planes.**

If we now use u and v for the distances of the object and image from the respective principal points, the formula **(28·61)** will also apply to the present case. It will be seen that **(28·61)** is satisfied by $u = v = 0$; so that the principal points are **conjugate points,** i.e. when the object is in the first principal point the image is in the second one. Further, the principal planes are planes of unit (lateral) magnification, since by **(28·66)**

$$\frac{y'}{y} = -\frac{f_1}{x},$$

or

$$\frac{y'}{y} = -\frac{f_1}{(u - f_1)};$$

and when $u = 0$ this becomes

$$\frac{y'}{y} = +1.$$

We may notice in passing that the longitudinal magnification is given by

$$\frac{dv}{du} = -\frac{f_1}{f_2}\frac{v^2}{u^2},$$

and in the immediate neighbourhood of the principal points approaches the limiting value

$$-f_2/f_1.$$

Equation **(28·61)** may also be satisfied by

$$u = v = f_1 + f_2.$$

The pair of conjugate points which are thus determined are called **nodal points.** Since the two nodal points are situated at equal distances from the respective principal points, the distance between them is the same as that between the principal points. A ray of light, in the object space, passing through the first nodal point N_I (Fig. 28·63) will cut the first principal plane in some point, A, or will do so when produced. The corresponding emerging ray in the image space must pass through the second nodal point, N_{II}, since the image of a point in N_I is situated in N_{II}. It will also cross the second principal plane at some point B; and the property of the principal planes, of

unit lateral magnification, requires that $P_IA = P_{II}B$. Finally $(N_IP_I) = N_{II}P_{II})$, since each is equal to $f_1 + f_2$. It follows therefore that the two rays are parallel to one another.
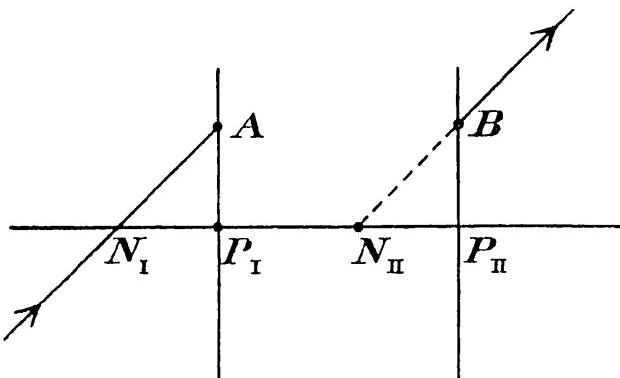
FIG. 28·63

It is of interest to inquire about the circumstances under which the nodal points coincide with the respective principal points. The condition for this is evidently

$$f_1 + f_2 = 0,$$

and therefore by (28·681) and (28·682)

$$f^a{}_2 f^b{}_2 = f^a{}_1 f^b{}_1,$$

and by (28·611) this is equivalent to

$$\left(\frac{n_2 r_a/n_1}{n_2/n_1 - 1}\right) \times \left(\frac{n_3 r_b/n_2}{n_3/n_2 - 1}\right) = \left(\frac{r_a}{n_2/n_1 - 1}\right)\left(\frac{r_b}{n_3/n_2 - 1}\right)$$

or

$$\frac{n_2}{n_1}\cdot\frac{n_3}{n_2} = 1,$$

and therefore

$$n_1 = n_3.$$

In words, the refractive indices of the image and object spaces must be equal. When this is the case then, the nodal points coincide with the respective principal points, and the focal lengths of the system are now numerically equal, but have opposite signs.

The method just described can easily be extended to a system consisting of any number of co-axial spherical refracting surfaces.

§ 28·7. Angular Magnification associated with Beams of Parallel Rays

When the incident and emergent beams of light of some optical system, such, for example, as a telescope or the prism of a spectrometer, are made up of parallel rays—beams of light

originating in the same object point are here meant—and when the media on both sides of the apparatus are optically identical, i.e. have equal refractive indices, the **angular magnification** of the apparatus is equal (for small angular dimensions of the object) to the quotient of the breadth of the incident beam by that of the emergent beam. In Fig. 28·7 the apparatus, whatever it may be, is indicated by the shaded portion $AA'B'B$. The angle between the directions of the incident beams from the extreme points (in the plane of the paper) is represented by θ, while θ' represents the angle between the directions of the corresponding emergent beams. Ax is a plane perpendicular to one of the incident beams, while $A'x'$ is a plane perpendicular to the corresponding emergent beam. Ay and $A'y'$ refer in a similar way to the beams associated with the other extreme point. The length of the optical path through the apparatus from A
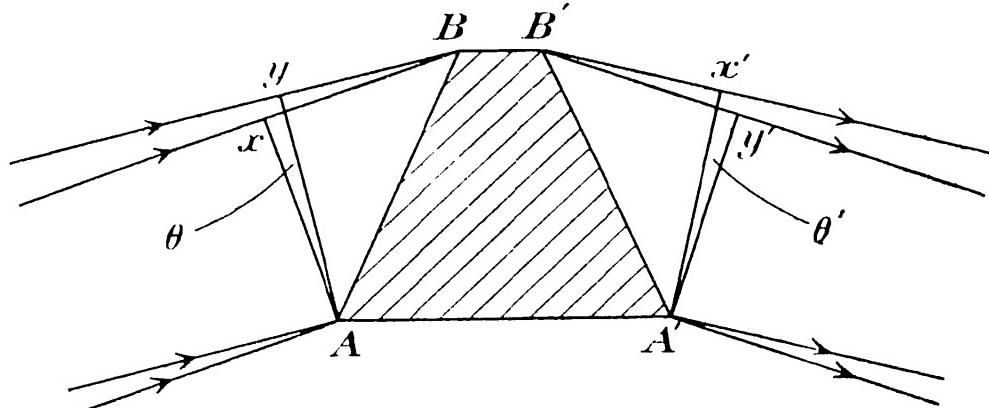
Fig. 28·7

to A' is equal to that from x to x', and is also equal to that from y to y'. Hence the optical path (xy) is equal to the path $(x'y')$. Consequently the actual distances xy and $x'y'$ are equal when the optical medium on both sides of the apparatus is the same. Now

$$\theta' = (x'y')/B_e,$$

and 　　　　　　　$$\theta = (xy)/B_i,$$

where B_e and B_i mean the breadths of the emergent and incident beams respectively. Hence the magnification is

$$\frac{\theta'}{\theta} = B_i/B_e.$$

§ 28·8. Spherical Aberration—Sine Condition—Aplanatic Systems

Let us consider an optical system such as a lens, with axial symmetry. The points O and A are situated in a plane per-

pendicular to the axis of the system (Fig. 28·8). We are going
to study the formation of images of these points when the rays
of light entering the lens system are not confined to directions
making small angles with the axis. In the last section we have

FIG. 28·8

seen that there is one
axial image point cor-
responding to each axial
object point when the
rays of light passing
through the system all
make small angles with
the axis. In general—
and a single spherical
refracting surface fur-
nishes a good illustration
—a ray from O making
a big angle, θ, with the
axis will, after passing
through the system, cut

the axis again (or will if produced backwards) at a point which
is separated by an appreciable distance from that where it is
cut by a paraxial ray. This defect is called **aberration** and,
when due to spherical refracting surfaces, **spherical aberration.**
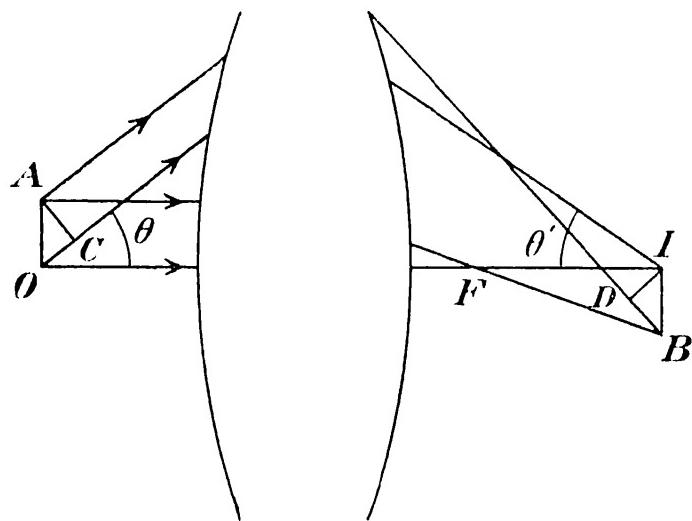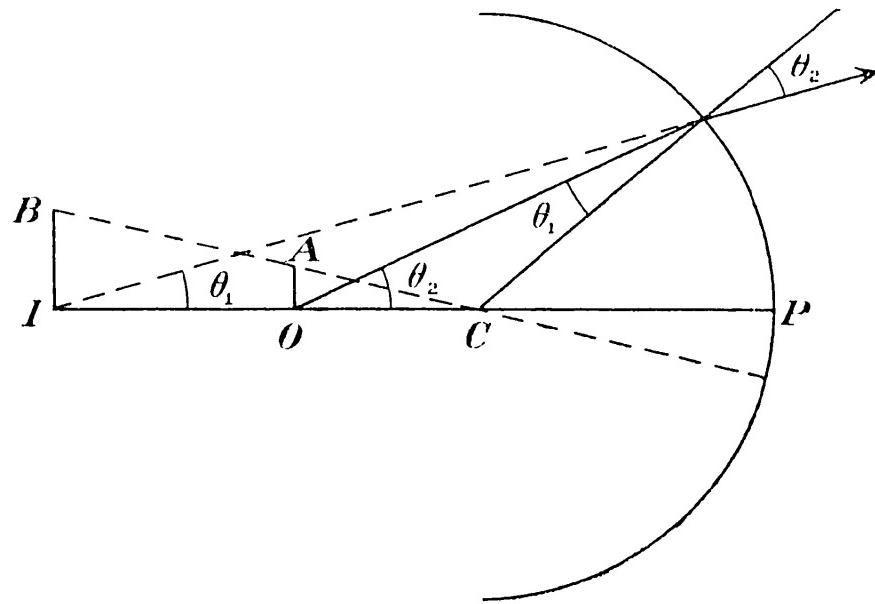
FIG. 28·81

In the case of a *single* spherical refracting surface there is
one place for a point object, which is free from aberration. This
is the point, O, in Fig. 28·81. If r be the radius of the sphere
and a and b the respective absolute distances of O and I,

the object and image points, from the centre, C, of the sphere ; then

$$a/r = n_2/n_1, \qquad \qquad \textbf{(28·8)}$$

and

$$ab = r^2. \qquad \qquad \textbf{(28·81)}$$

It will easily be seen that all rays from O, after refraction at the spherical surface, will pass through I if produced backwards. This will be the case for all values of the angle of incidence, θ_1. It will be noticed that the angle of refraction, θ_2, is identical with the angle between the incident ray, OM, and the axis, OP; while the angle of incidence, θ_1, is identical with the angle between the refracted ray, IM, and the axis.

Even when the optical system is so constituted (as in the illustration just given) that a point object at some point, O, on the axis is reproduced as a point image by pencils of rays making all angles small or great with the axis ; in other words, when there is no spherical aberration, it is not in general true that there is an image point, B, corresponding to such an object point as A in Fig. 28·8 ; A and B being points near the axis and in planes perpendicular to the axis through O and I respectively. We shall now investigate the further condition to be satisfied in order that this may be the case. To begin with, we agree that all the rays from O in Fig. 28·8 pass through a point, I, or, failing that, all the rays in the image space pass through a point, I, if produced backwards.

Consider now two rays from O to I, one, the axial ray, and the other a ray making any angle, θ, and a corresponding angle θ', with the axis, as shown in Fig. 28·8. Both paths from O to I have of course the same optical length. The two corresponding rays from the point, A, one parallel to the axis, and the other making the angle θ with it, have to pass through a point B in a plane through I perpendicular to the axis. OA and IB we are supposing to be (first order) small quantities. Now since OA is perpendicular to the axis, the optical length of the axial ray from O to I and of the ray from A to B which starts parallel to the axis, must be equal, i.e. can differ only by a second order small quantity. In fact as far as the focus, F, they are exactly equal, and the remaining portions, FI and FB, can only differ by a second order small quantity. It follows now that the optical length of the two parallel rays, one from O and the other from A making an angle, θ, with the axis must also have equal optical lengths. Hence

$$n(OC) = n'(DB).$$

Now
$$(OC) = y \sin \theta,$$
$$(DB) = y' \sin \theta',$$

therefore

$$n(OC) = ny \sin \theta,$$
$$n'(DB) = n'y' \sin \theta',$$

and so we find the condition :

$$\frac{ny}{n'y'} = \frac{\sin \theta'}{\sin \theta} . \qquad \qquad \textbf{(28·82)}$$

This is known as the **sine condition.** In general there is only *one* point, O, and its associated image point, I, for which the sine condition can be satisfied. It may be noted that **(28·82)** is in agreement with **(28·65),** since in the latter formula θ and θ' are very small angles. It is easy to show that the sine condition holds in the case of a single refracting spherical surface (Fig. 28·81) when the object is placed in the position, O, which is indicated. Let OA be the small object ; OA being perpendicular to the axis IOC. The image of A is at some point B, which, if we neglect second order small quantities, will be in the plane through I perpendicular to the axis. That this is the case can be seen by constructing a new axis through C and A.

Now by **(28·63)** the ratio

$$\frac{y'}{y} = \frac{IB}{OA} = \frac{n_1 v}{n_2 u}.$$

Therefore

$$\frac{y'}{y} = \frac{n_1(b + r)}{n_2(a + r)},$$

and by **(28·8)** and **(28·81)**

$$\frac{y'}{y} = \frac{n_1(n_1/n_2 + 1)}{n_2(n_2/n_1 + 1)},$$

or

$$\frac{y'}{y} = \frac{n_1^2}{n_2^2}.$$

Consequently

$$\frac{n_2 y'}{n_1 y} = \frac{n_1}{n_2} = \frac{\sin \theta_2}{\sin \theta_1}. \qquad \qquad \textbf{(28·83)}$$

The angles θ_1 and θ_2 here mean the angles of incidence and refraction respectively ; but θ_2 is also the angle between the incident ray and the axis, while θ_1 is that between the refracted ray and the axis as already explained. When this is remembered it will be seen that **(28·83)** is identical with the sine condition.

Systems for which points such as O and I exist, which are free from aberration and also satisfy the sine condition, are called by Abbe **aplanatic systems.**

The theory outlined in this section has a very important application in connexion with the construction of microscope objectives.

§ 28·9. CHROMATIC ABERRATION

So far we have supposed the optical systems to be traversed by light of one wave-length, and therefore of one refractive index for a given medium—by monochromatic light in fact. Most optical instruments are used to form images of objects with white light, which is a mixture of light of all wave-lengths within the range extending roughly from 4000 $A.U.$ to 8000 $A.U.$[1]

A lens system forms an image of the object for every wave-length, and in general these images are in different situations and of different magnifications. The appearance of such a collection of images of different colours and sizes and in slightly different positions, is that of a single rather rough image very much coloured, especially at the edges. This kind of defect is called **chromatic aberration.** Lens systems can be designed so that it is greatly minimized. The detailed treatment of chromatic aberration and the methods by which it may be reduced are outside the scope of this volume.

BIBLIOGRAPHY

NEWTON: Opticks. (1704.)

P. DRUDE: Lehrbuch der Optik. (Hirzel, Leipzig.) Chapters I to V deal with geometrical optics.

E. T. WHITTAKER: The Theory of Optical Instruments. (Cambridge Tracts.)

---

[1] An Angstrom unit of length ($A.U.$) is 10^{-8} centimetre.

CHAPTER XII

PHYSICAL OPTICS—PHENOMENA OF INTERFERENCE AND DIFFRACTION

§ 29. PRINCIPLE OF SUPERPOSITION

THE partial differential equations which represent waves are of the linear type ; i.e. the dependent variable and its differential quotients do not appear in them in higher powers than the first. A consequence of this, often pointed out already, is that the sums of particular solutions are also solutions of such equations. The most complicated state of affairs represented by such an equation can be described by Fourier's methods as a sum of functions (proper functions—*Eigenfunktionen*) such as sine and cosine functions, Bessel's functions, spherical harmonic functions and so on. An important special case is that of two plane harmonic waves which pass over and occupy a region common to both. Each of them separately is described by a particular simple harmonic solution of the wave equation, and in the region common to both the sum of these is the appropriate solution. This is the **principle of superposition.**

The undulatory theory of light necessarily leads to the conclusion that light spreads round obstacles and penetrates regions which a naïve corpuscular theory, involving a strictly linear propagation, declares to be regions of shadow. We have already (**§§ 27·1 and 28·2**) pointed out that when the wave-length of the light is sufficiently short linear propagation without spreading is practically a fact. We term the spreading which is observable when the wave-length is not too short (compared with the dimensions of apertures, lenses, etc.) **diffraction.** Another phenomenon called **interference** is observed where coherent [1] beams of light meet or overlap. This phenomenon is explained by the **principle of superposition** and is well illustrated by the classical experiment of Thomas Young, of which the experiments with Fresnel's mirrors and bi-prism are more elaborate variants.

[1] Cf. **§ 30·1.**

§ 30. YOUNG'S EXPERIMENT

In the classical experiment of Thomas Young a source of monochromatic light, *s* in Fig. 30, falls on a screen with two small apertures, *a* and *b*, near to one another. The waves which emerge from them proceed onwards to a screen, *P*. When the apertures are of the same size and equally distant from *s*, we may regard them as identical point sources of light. The displacements which they would, separately, produce at a point, *M*, on the screen may be represented by
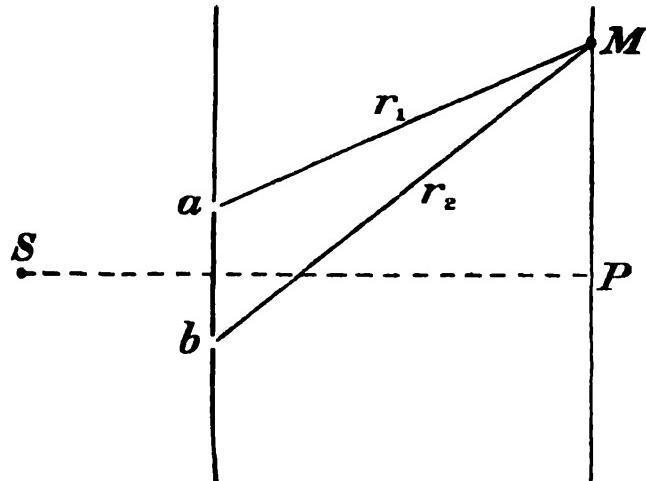
FIG. 30

$$s_1 = \frac{a}{r_1} \cos 2\pi\left\{\frac{t}{\tau} - \frac{r_1}{\lambda}\right\},$$

$$s_2 = \frac{a}{r_2} \cos 2\pi\left\{\frac{t}{\tau} - \frac{r_2}{\lambda}\right\} ;$$

the amplitudes at *M* being (cf. **§ 9·2**) inversely proportional to the distances r_1 and r_2. The principle of superposition asserts that the resultant displacement at *M* is

$$s_1 + s_2 = \frac{a}{r_1} \cos 2\pi\left(\frac{t}{\tau} - \frac{r_1}{\lambda}\right) + \frac{a}{r_2} \cos 2\pi\left(\frac{t}{\tau} - \frac{r_2}{\lambda}\right).$$

Now the cosine terms will be identical for all points, *M*, where $r_2 - r_1 = n\lambda$, *n* being an integer. At such points

$$s_1 + s_2 = \left(\frac{a}{r_1} + \frac{a}{r_2}\right) \cos 2\pi \left(\frac{t}{\tau} - \frac{r_1}{\lambda}\right),$$

and the resulting amplitude is the sum of the amplitudes of the two wave trains. On the other hand, at points where $r_2 - r_1 = (n + \frac{1}{2})\lambda$ we have

$$s_1 + s_2 = \left(\frac{a}{r_1} - \frac{a}{r_2}\right) \cos 2\pi \left(\frac{t}{\tau} - \frac{r_1}{\lambda}\right),$$

the resultant amplitude being equal to the difference of the individual amplitudes. The illumination of the screen therefore exhibits a set of bright fringes (interference fringes) with dark

spaces between them. This device of Young is typical of a group of **interferometers** which function by dividing the wavefront from some source like s in Fig. 30.

§ 30·1. COHERENCE

The description of interference, and in particular of Young's experiment, in the last section only takes account of the bare essentials. Any laboratory source of light such as s in Fig. 30 consists of an enormous number of elementary emitting systems (atoms or molecules), and we always have, in fact, innumerable pairs of interfering wave trains, aM and bM, which are initiated or stopped in an irregular and quite incalculable way. Since however the members of each pair start from one and the same emitting atom, the phase difference between the waves emerging at a and b will always remain the same, and in consequence the intensity of illumination at any point, M, on the screen will be completely determined by the difference of the optical lengths of r_1 and r_2 in the way already explained. If instead of a single original source at s we had two original sources, one at a and the other at b, we should not then be able to regard the emissions from them as made up of pairs of harmonic wave trains exactly alike and having a constant phase difference at a and b. The emission from an atom in a would have no correlation whatever with that of a similar atom in b, and the orderly interference described in **§ 30** could not occur. When, as in the experiments of Young or Fresnel, two beams of light are capable of interfering with one another, they are said to be **coherent**.

§ 30·2. PARALLEL BEAM AND RECTANGULAR APERTURE

We are now going to study the case of a parallel beam of monochromatic light which for the present we may suppose to originate in a point source, and to have been made parallel by a suitably placed lens or lens system (collimator). Let us suppose it to be limited by the parallel sides of a slit-shaped aperture on which it falls perpendicularly. After passing the aperture the beam will, in the limiting case of very short wavelength, remain strictly parallel, and if it falls on a suitably placed lens a real point image will be formed at the principal focus of the lens. The amplitude of the beam, i.e. the maximum value of the electric displacement in it, is proportional to the square root of the intensity of the beam, that is to the amount of energy transported through the unit area per unit time, and

it is convenient, for purposes like the present one, simply to identify it with the root of the intensity. Since the amplitude of waves proceeding from a point source is inversely proportional to the distance from the source (§ 9·2), the amplitude at the point source itself is infinite unless we modify our conception of a point source in a similar way to that adopted for a point charge of electricity. We must regard it, not as a mathematical point, but a small region over which the amplitude has a finite value.

If we have a source of light consisting of a continuous line of point sources (parallel to the sides of the aperture), such as is exemplified by the illuminated slit of a spectrometer, then— if we continue to imagine the limiting case of extremely short waves—we shall have associated with each point of the source (slit) a parallel beam and a corresponding point image ; all these images making up a line which is an image of the source or slit.

We are interested in the case where the wave-length is long enough for diffraction phenomena to occur. These can only be dealt with completely by applying Huygens' principle in its rigorous form. The principle will be demonstrated in § 31·8, and meanwhile we shall attack the

Fig. 30·2

problem in a more elementary way, leaving till later the justification of any assumptions we may make.

It should be noted that the parts of a beam *originating in the same point source* are coherent, and can therefore interfere with one another ; but beams, or portions of them, which originate in different point sources are not coherent, and do not interfere with one another. A consequence of this is that any spreading in directions parallel to the slit (i.e. to the edges of the rectangular aperture) will merely elongate slightly the image of the slit. Things are quite different with lateral spreading. We shall regard the light which has passed the aperture, and all of which originates in the same slit point, as made up of an infinite number of parallel beams which make angles, θ (in a plane perpendicular to the aperture), as shown in Fig. 30·2. The slit is so placed, we are supposing, that all points along AB are points of equal phase. AC is a perpendicular section of a beam making an angle, θ, with AB. Let us suppose this
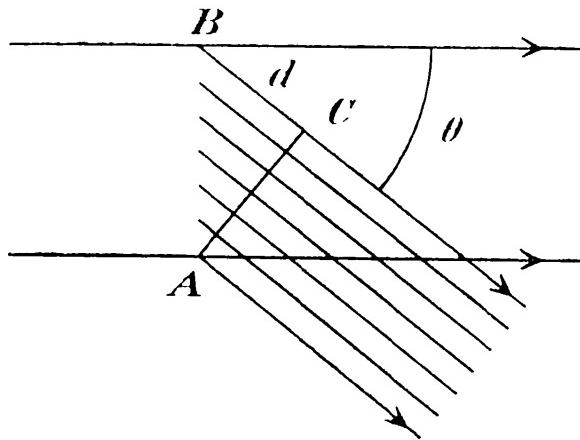
beam to be divided into a large (infinite) number of parts by planes parallel to its direction and to the sides of the aperture. Each of these parts would, by itself, give rise to some amplitude, a, in the corresponding image point produced by a lens, the axis of which is coincident with that of the beam. The application of the principle of superposition to all these portions of the beam yields for the displacement, s, in the image point the sum

$$s = a \cos \omega t + a \cos (\omega t + \delta) + \ldots + a \cos (\omega t + (n-1)\delta), \quad \textbf{(30·2)}$$

where n is the number of divisions, and δ is the increment in phase as we proceed from C to A, ωt being the phase in C. The sum in **(30·2)** is easily shown to be

$$s = \frac{a \cos \left\{ \omega t + \dfrac{(n-1)}{2}\delta \right\} \sin \dfrac{n\delta}{2}}{\sin \dfrac{\delta}{2}}$$

or, if $na = A$,

$$s = \frac{A \sin \dfrac{n\delta}{2} \cos \left(\omega t + \dfrac{n-1}{2}\delta \right)}{n \sin \dfrac{\delta}{2}}. \quad \ldots \quad \textbf{(30·201)}$$

If we represent $n\delta$, the phase difference between the extreme sides of the beam, by ϕ and let n become infinitely large while δ in consequence converges to zero, **(30·201)** becomes

$$s = \left(A \frac{\sin \phi/2}{\phi/2} \right) \cos (\omega t + \phi/2). \quad \ldots \quad \textbf{(30·21)}$$

Now $$\phi = 2\pi d/\lambda,$$

where d is the distance BC, and

$$d/B = \theta,$$

if B is the breadth of the aperture, and if we confine our attention to small values of θ. Therefore

$$\frac{\phi}{2} = \frac{\pi B \theta}{\lambda};$$

so that in the direction, θ, the amplitude is

$$A \frac{\sin \dfrac{\pi B \theta}{\lambda}}{\dfrac{\pi B \theta}{\lambda}} \quad . \quad . \quad . \quad . \quad . \quad \textbf{(30·22)}$$

It will consequently vanish in directions for which

$$\frac{\pi B \theta}{\lambda} = \nu \pi, \qquad . \quad . \quad . \quad . \quad . \quad (30 \cdot 23)$$

where ν is any integer other than zero. In place therefore of an ideal image of the slit we shall have a number of bright fringes with dark ones in between; the first dark fringe being separated from the central bright one by the angle which corresponds to $\nu = 1$ or $\nu = -1$, or the angle

$$\theta = \lambda/B. \quad . \quad . \quad . \quad . \quad . \quad (30 \cdot 24)$$

If we denote $\pi B \theta / \lambda$ by α, so that the amplitude is $A \sin \alpha / \alpha$, and if we assume A to be independent of θ—a fairly safe assumption so long as θ is small—we may obtain the positions of maximum brightness from

$$\frac{d}{d\alpha} \left(\frac{\sin \alpha}{\alpha} \right) = 0,$$

or

$$\tan \alpha = \alpha. \quad . \quad . \quad . \quad . \quad (30 \cdot 25)$$

The corresponding values of α, and therefore also of θ, can easily be found graphically. All we have to do is to plot $\tan \alpha$ against α and draw a straight line through the origin, making an angle of 45° with either axis. Its intersections with the curve will give the required values of α.[1]

If in any spectroscopic device the emerging parallel beams due to spectral lines of wave-length λ and $\lambda + d\lambda$ are separated by an angle $d\theta$ such that

$$d\theta = \lambda/B, \quad . \quad . \quad . \quad . \quad (30 \cdot 26)$$

(see **30·24**), the central fringe due to one of them will be formed just where the first dark fringe due to the other is produced, and in these circumstances, as Lord Rayleigh showed, the two lines will appear resolved. They will, of course, appear resolved even when the separation is slightly smaller than this; but this particular separation has been adopted as the conventional criterion for the resolution of neighbouring lines.

If we divide both sides of **(30·26)** by $d\lambda$ we have

$$\frac{d\theta}{d\lambda} = \frac{\lambda}{B d\lambda}.$$

[1] Cf. Schuster and Nicholson's *Optics*, p. 104.

The ratio $\lambda/d\lambda$ is called the **resolving power** of the apparatus and $d\theta/d\lambda$ the **dispersion.** We therefore find

$$BD = R, \quad . \quad . \quad . \quad . \quad . \quad . \quad \textbf{(30·27)}$$

where B is the breadth of the emerging beam, D is the dispersion and R the resolving power.

§ 30·3. Resolving Power of a Prism

Consider a parallel beam of light of various wave-lengths incident on one face AC of a prism, so as to be perpendicular to its refracting edge, C (Fig. 30·3). The extreme sides of the beam traverse through the prism the respective distances zero and $AB = t$. Let xA be a plane perpendicular to the incident beam,
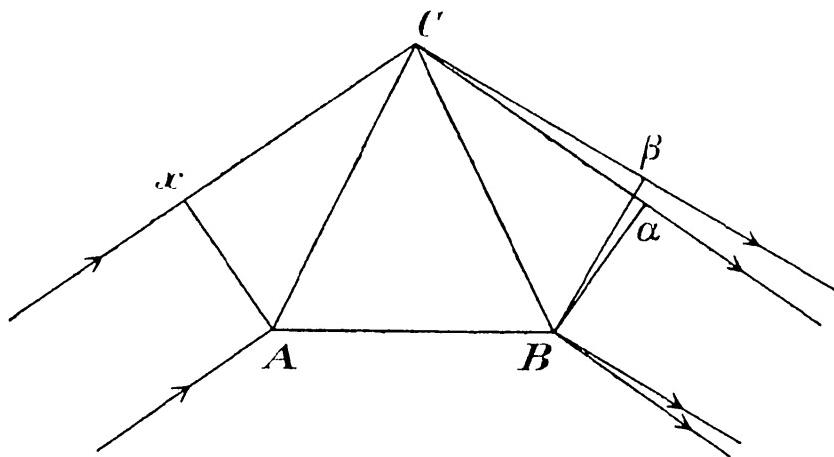
Fig. 30·3

αB a plane perpendicular to the emergent beam of wave-length λ, βB that perpendicular to the emergent beam of some wave-length $\lambda + d\lambda$. The absolute value of the angle, $d\theta$, between the two emergent beams is

$$d\theta = (\alpha\beta)/B, \quad . \quad . \quad . \quad . \quad \textbf{(30·3)}$$

where B is the breadth of an emergent beam in the neighbourhood of λ. Now the optical distance along AB is equal to $xC + C\alpha$, or

$$\mu t = xC + C\alpha,$$

where μ is the refractive index of the glass for the wave-length λ. Similarly,

$$(\mu + d\mu)t = xC + C\beta,$$

where $d\mu$ corresponds to $d\lambda$. Hence

$$d\mu.t = - (\alpha\beta). \quad . \quad . \quad . \quad . \quad \textbf{(30·31)}$$

The sign adopted corresponds to the fact that μ and θ diminish with increasing λ and the figure has been drawn in accordance.

On combining the equations **(30·3)** and **(30·31)** and dividing by $d\lambda$, we get

$$R = B\frac{d\theta}{d\lambda} = -\frac{d\mu}{d\lambda}.t, \quad . \quad . \quad . \quad . \quad \textbf{(30·32)}$$

which is the resolving power of the prism. It is easy to see that when the incident beam does not extend to the refracting edge, C, of the prism, the resolving power is equal to $-d\mu/d\lambda$ multiplied by the difference, $t_2 - t_1$, of the distances travelled through the prism by the extreme sides of the beam.

§ 30·4. PARALLEL BEAM AND CIRCULAR APERTURE

We now turn to the case where the aperture is circular and the parallel beam of monochromatic light originates in an axial point source. As in **§ 30·2**, we imagine the beam divided into an infinite number of elementary portions; but this time by circles situated in the plane of the aperture and having their centres coincident with that of the aperture, and by intersecting radial lines. We may represent the contribution of each elementary area to the total amplitude produced in the focal region of the collecting lens by

$$a\rho \, d\rho \, d\phi,$$

where ρ and $\rho + d\rho$ are the radii of the circles bounding the elementary area, and $d\phi$ is the azimuthal angle between the two bounding radii. In Fig. 30·4 the aperture is represented as perpendicular to the plane of the paper by the line AB. Just as in **§ 30·2**, we shall regard the light which has passed the aperture as consisting of an infinite number of beams of parallel light making various angles, θ, with the axial direction. The plane AC is perpendicular to a beam travelling in the direction, θ, and x represents the position of one of the elementary areas projected on the plane of the paper. Its perpendicular distance from the plane AC is represented by d. The phase difference associated with d is

FIG. 30·4

$$\delta = 2\pi d/\lambda.$$

Now

$$d = \rho \cos \phi \sin \theta + R \sin \theta,$$

if we measure the azimuthal angle, ϕ, from OB, and if R be the
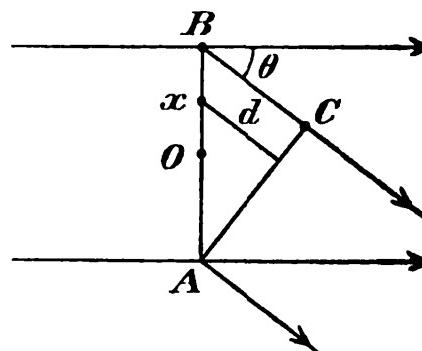
radius of the aperture.   We find therefore for the equation which corresponds to **(30·2)**

$$s = a \int_0^R \int_0^{2\pi} \rho d\rho d\phi \, \cos \, (\omega t + \delta),$$

or

$$s = a \int_0^R \int_0^{2\pi} \rho d\rho d\phi \, \cos \, \left\{ \omega t + \frac{2\pi}{\lambda}(R \sin \theta + \rho \cos \phi \sin \theta) \right\}.. \qquad \textbf{(30·4)}$$

We may only regard a as constant for small values of θ and for these the expression for s becomes

$$s = a \int_0^R \int_0^{2\pi} \rho d\rho d\phi \, \cos \, \left\{ \omega t + \frac{2\pi\theta}{\lambda}(R + \rho \cos \phi) \right\}.$$

This is easily seen to be equivalent to

$$s = a \cos \left(\omega t + \frac{2\pi\theta R}{\lambda} \right) \int_0^R \int_0^{2\pi} \rho d\rho d\phi \, \cos \left(\frac{2\pi\theta\rho}{\lambda} \cos \phi \right),$$

since the integral involving the *sine* is equal to zero.   Hence the resultant amplitude in the direction θ is

$$a \int_0^R \int_0^{2\pi} \rho d\rho d\phi \, \cos \left(\frac{2\pi\theta\rho}{\lambda} \cos \phi \right). \qquad . \qquad . \qquad \textbf{(30·41)}$$

We now expand the integrand in ascending powers of $2\pi\theta\rho \cos \phi/\lambda$, and integrate term by term.   We thus obtain for the amplitude,

$$\frac{\pi D^2 a}{4} \left\{ 1 - \frac{x^2}{2(1!)^2} + \frac{x^4}{3(2!)^2} - \frac{x^6}{4(3!)^2} + \cdots \right\}, \qquad . \qquad . \qquad \textbf{(30·42)}$$

where $\qquad\qquad\qquad x \equiv \pi\theta D/2\lambda,$

and D ($\equiv 2R$) is the diameter of the aperture.   The series in **(30·42)** is

$$J_1(2x)/x,$$

i.e. the Bessel function, J_1, (of $2x$) divided by x.   It vanishes for an infinite number of positive real values of x, the smallest being

$$x = 1·92 \ldots$$

Instead therefore of a point image of the point source we get a

central bright spot surrounded by bright rings. The dark ring between the central bright spot and the first bright ring is associated with the value of θ corresponding to $x = 1\cdot92 \ldots$, or

$$\theta = 1\cdot22 \ \lambda/D.[1] \ . \ . \ . \ . \ . \quad (30\cdot43)$$

A star, which we may regard as a point source of light, appears in an astronomical telescope as such a system of rings.

§ 30·5. Resolving Power of Telescope

When the angular separation, θ, of two stars is just so great that the central bright spot of one falls in the first dark ring due to the other, it is then possible to distinguish that there are two ring systems, and thus to recognize that two stars are being observed. In this case

$$\theta = 1\cdot22 \ \lambda/D,$$

as we have already seen. The reciprocal of this angle is used as the conventional measure of the **resolving power** of the telescope. It is all the greater the shorter the effective wavelength λ, and the greater the diameter of the object glass of the telescope.

§ 30·6. The Stellar Interferometer

Let us suppose the object glass of a telescope to be covered by a screen ab (Fig. 30·6) with two small apertures or slits, c and d, situated on a line through its centre and equally distant from it. Let D be the distance apart of c and d, and let AB be the axis of the telescope, the principal focus of the objective being at B. Further, let C be any point on a line through B perpendicular to the axis, and L_1 and L_2 its respective distances from the apertures c and d. If we represent the distance CB by r, we have

Fig. 30·6

$$(D/2 + r)^2 + F^2 = L_1{}^2,$$
$$(D/2 - r)^2 + F^2 = L_2{}^2,$$
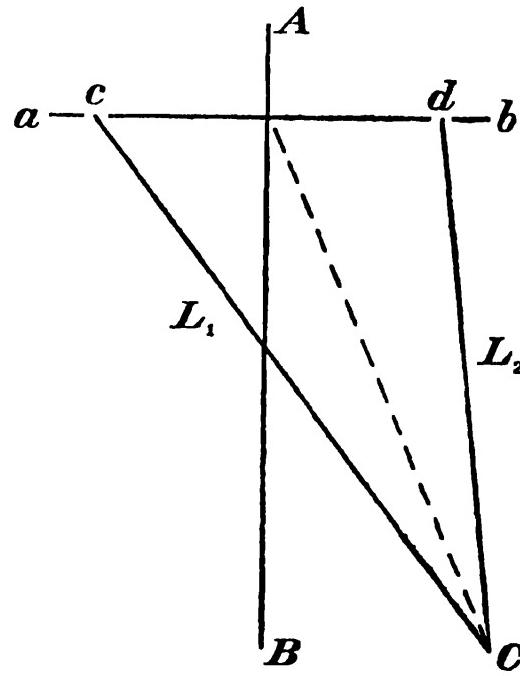
[1] G. B. Airy, *Camb. Phil. Trans.*, p. 283 (1834).

where F is the focal length of the objective. Hence, approximately,

$$r/F = (L_1 - L_2)/D,$$

since L_1, L_2 and F are not very different. If θ be the angle between AB and AC,

$$\theta = (L_1 - L_2)/D, \quad . \quad . \quad . \quad . \quad \textbf{(30·6)}$$

and bright fringes will therefore appear, as in Young's experiment, at those points, C, for which

$$\theta = n\lambda/D, \quad . \quad . \quad . \quad . \quad . \quad \textbf{(30·61)}$$

and the first dark place on each side of the central bright fringe will occupy the position

$$\theta = \lambda/2D. \quad . \quad . \quad . \quad . \quad \textbf{(30·62)}$$

The displacement at C is equal to

$$\frac{a}{L_1} \cos 2\pi\left\{\frac{t}{\tau} - \frac{L_1}{\lambda}\right\} + \frac{a}{L_2} \cos 2\pi\left\{\frac{t}{\tau} - \frac{L_2}{\lambda}\right\},$$

where a is the amplitude in either of the apertures c or d. Since L_1 and L_2 are nearly equal, this may be written

$$\frac{2a}{L} \cos 2\pi\left\{\frac{t}{\tau} - \frac{L}{\lambda}\right\} \cos 2\pi\left\{\frac{L_1 - L_2}{2\lambda}\right\}.$$

Consequently the amplitude at C is

$$\frac{2a}{F} \cos 2\pi\left\{\frac{L_1 - L_2}{2\lambda}\right\},$$

or

$$\frac{2a}{F} \cos \left(\frac{\pi D\theta}{\lambda}\right),$$

and the intensity

$$\frac{4a^2}{F^2} \cos^2 \left(\frac{\pi D\theta}{\lambda}\right). \quad . \quad . \quad . \quad . \quad \textbf{(30·63)}$$

If therefore the telescope be directed to a star, a set of bright fringes spaced according to **(30·61)** will be seen, and the intensity of the light will vary through the fringe system according to the law **(30·63)**. When the angular distance, θ, between two neighbouring and equally bright stars is given by the expression **(30·62)**, the central bright fringe due to the one falls in the middle of the first dark one of the other, and in accordance with **(30·63)** uniform illumination will result.

When the distance, D, can be adjusted, the fringes will be

widely separated when it is small and approach one another more and more closely as it is increased. Two very slightly separated stars will therefore give rise to visible fringes when D is not too great ; but when it is increased to the value corresponding to (30·62) they will completely vanish. The angular separation of the stars will then be given by (30·62). This kind of device can hardly be said to function as a telescope ; it is, in fact, an interferometer, and its maximum resolving power is

$$1/\theta = 2D/\lambda,$$

where D is the diameter of the objective ; while that of the unmodified telescope is

$$1/\theta = D/1\cdot22\ \lambda,$$

approximately only half as much.

§ 30·7. MICHELSON'S MAGNIFYING DEVICE

If the angular diameter of a star's disc were not so exceedingly minute as it is even in the case of the nearest stars, the method described above might be employed to measure it. It appears to have been suggested by Fizeau. Let us for the moment regard a star as a uniformly bright *linear* object parallel to the line cd of the apertures in the screen over the telescope objective. It is clear that, so long as the angle subtended at the telescope by this linear object is less than the angular separation of successive fringes due to a point source, fringes will be observed when the telescope is directed to it, and they will just disappear when the angle is equal to λ/D. Actually the star is like a circular disc, so that more light comes from the central portion than from the sides, with the result that the fringes do not disappear until the angular diameter slightly exceeds λ/D. The angular diameter of a star should therefore be approximately determinable by increasing the separation of c and d until the fringes just disappear. It turns out that the aperture of the largest astronomical telescope is not nearly wide enough for this purpose—an aperture of 20 to 40 feet or even more being necessary or desirable to render possible the measurement of the angular diameters of very small or very remote stars. The difficulty has been surmounted by Michelson by the ingenious magnifying device illustrated in Fig. 30·7. In this device there are four mirrors. Two of them, M_1 and M_4, are mounted on a long girder in front of the telescope objective. The light from the star is reflected from these to two inner mirrors, M_2 and M_3, from which the beams enter the objective, O. We may regard

these inner mirrors as playing the part of the apertures c and d of Fig. 30·6. When the proper adjustments have been made to secure that the beams overlap in the focal plane, F, a diffraction image, S, of the star disc is seen crossed by fine fringes A'B'C'D'. The arrangement produces a virtual image of the star, magnified in angular diameter by D/D' (cf. § 28·7), D being the distance between the outer mirrors and D' that between the inner ones.
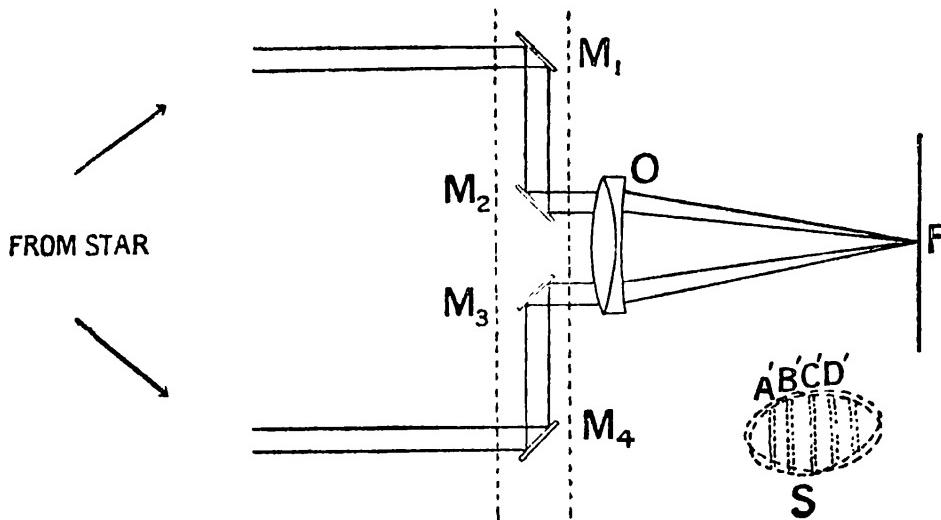
FIG. 30·7

The angular diameter *of this image* is λ/D', assuming the outer mirrors, which are movable, have been shifted just sufficiently far outwards to cause the fringes to vanish ; or, strictly speaking, the angular diameter slightly exceeds this. But this measured angular diameter is D/D' times as great as that of the star. Hence the angular diameter of the star is a little greater than

$$\theta = \frac{\lambda}{D'} \cdot \frac{D'}{D} = \frac{\lambda}{D}.$$

We have tacitly assumed the light from the star to be monochromatic, whereas in fact it is, roughly speaking, white light. It is found in practice that a moderately narrow range of wavelengths is effective in producing the fringes, and it is this wavelength which should be substituted in the formula. Michelson and his collaborators found that the fringe system due to the star Betelgeuse in the constellation Orion (α *Orionis*) disappeared when the outer mirrors were separated by 121 inches ($D = 121$ inches). Taking the effective wave-length to be 5,750 $A.U.$ (an Angstrom unit $= 10^{-8}$ cm.), we find for θ,

$$\theta = \frac{5{,}750 \times 10^{-8}}{121 \times 2 \cdot 54} = \cdot 0386 \text{ seconds.}$$

This does not take into account the influence of the circular shape of the star's disc mentioned above. The angle θ is more accurately given by adding about 1/5 which makes $\theta = \cdot046$ seconds.

§ 30·8. THE PLANE GRATING

The transmission type of plane grating consists of a plate of transparent material on which equally spaced parallel lines have been ruled. It is in effect equivalent to a large number of very narrow parallel slits. If we regard it as taking the place of the rectangular aperture in **§ 30·2** we may apply to it the formula (**30·201**), in which, however, n is now a finite (though a large) integer. Let us write it in the form :

$$ s = \frac{a \sin \dfrac{n\delta}{2}}{\sin \dfrac{\delta}{2}} \cos \left(\omega t + \frac{n-1}{2} \delta \right), \quad . \quad . \quad (30\cdot8) $$

in which a means the amplitude associated with a single line of the grating. We are concerned with the dependence of the amplitude,

$$ a \sin \frac{n\delta}{2} \bigg/ \sin \frac{\delta}{2}, \quad . \quad . \quad . \quad . \quad (30\cdot81) $$

on the angle, θ (Fig. 30·8 (a)). Obviously if δ has the value 0, the amplitude is equal to

$$ na = A. $$

The substitution of $N\pi + \varepsilon$, where N is an integer, for $\delta/2$, gives

$$ a \frac{\cos \cdot nN\pi \sin n\varepsilon}{\cos N\pi \sin \varepsilon}. $$

This approximates to

$$ na = A $$

as ε approaches zero, if we ignore the sign, which may be taken care of by introducing a π into the phase. Now δ, the phase step from one line to the next (see Fig. 30·8 (a)), is equal to

$$ \delta = 2\pi d/\lambda, $$

where d is the distance,

$$ e \sin \theta, $$

e being the breadth of one line. Consequently when $\delta/2 = N\pi$, we find

$$d = N\lambda,$$

and hence

$$N\lambda = e \sin \theta. \quad . \quad . \quad . \quad . \quad . \quad (30\text{·}82)$$

For these values of θ therefore an image of the slit from which the light proceeds is produced, assuming, of course, that the light is brought to a focus by a suitably placed converging lens. For values of θ between those indicated by (30·82) there will be a succession of maxima of small amplitude (when n is great), as an examination of (30·8) indicates. The image corresponding to the value N is termed an Nth order image, and with light of several or many wavelengths there will be a large number of images of the slit, one for each wave-length for a given N. These images constitute the Nth order spectrum.
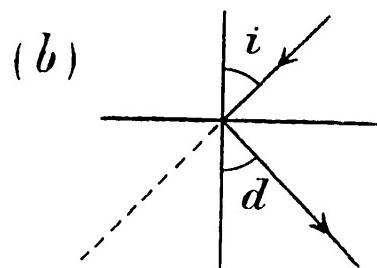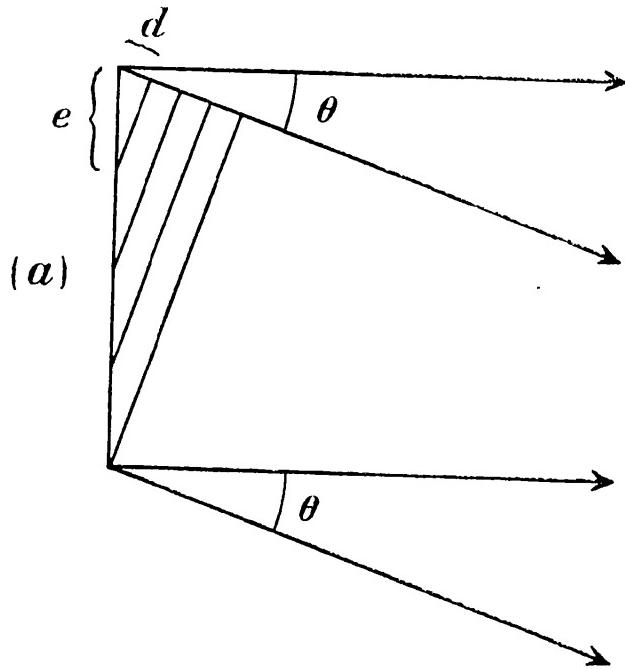
When the incident beam is incident at an angle i (Fig. 30·8 (b)), instead of being perpendicular as we have been supposing, then we have for the Nth order image

$$N\lambda = e \sin i + e \sin d,$$

or

$$N\lambda = 2e \sin \frac{i + d}{2} \cos \frac{i - d}{2}.$$

Now $D(= i + d)$ may be termed the angle of deviation. It is seen that this has its minimum value when $i = d$, and when this condition holds

$$N\lambda = 2e \sin \frac{D}{2}. \quad . \quad . \quad . \quad . \quad . \quad (30\text{·}83)$$

FIG. 30·8

§ 30·9. RESOLVING POWER OF GRATING

By differentiating **(30·82)** we get

$$\frac{d\theta}{d\lambda} = \frac{N}{e \cos \theta}.$$

This is the dispersion in the Nth order region. The breadth of the corresponding emergent beam is

$$B = ne \cos \theta.$$

Hence by **(30·27)** the corresponding resolving power is Nn, i.e. it is equal to the product of the number of grating lines over which the beam extends and the order of the spectrum.

§ 31. THE REFLECTING GRATING

The American physicist Rowland constructed reflecting gratings which were immensely superior to any wave-length measuring devices in existence before his time. The diamond point used to rule the lines on the speculum metal was displaced from line to line by a very accurately cut screw, so that great accuracy in their spacing was achieved. They were ruled on a concave surface of large radius. It was thus possible to dispense entirely with lenses, and their attendant disadvantage due to absorption. For a description of the concave grating the reader is referred to R. W. Wood's *Physical Optics* and other optical treatises mentioned at the end of this chapter.

§ 31·1. THE ECHELON GRATING

We have seen that the resolving power of a grating is equal to Nn, where n is the number of grating lines over which the beam spreads, and N is the order of the spectrum. In an ordinary grating the order is severely limited. It cannot exceed the number of times the wave-length of the light is contained in the interval between two consecutive lines, and this interval is necessarily very small in order that the grating lines may not spread over too great an area, and that they may not consequently necessitate the use of beams which are impossibly broad. The number of lines on a grating is also limited by the fact that the diamond point which rules them gradually becomes worn. It is essential that this wearing of the diamond point should be negligible during the ruling of the grating, in order that all the lines may be alike. Michelson devised a way of securing

enormously high resolving power by using a number of parallel plates arranged like the steps of a ladder (Fig. 31·1). This echelon, as it is called, is placed in the parallel beam between the collimator and the telescope of a prism spectrometer. The parallel beam of nearly monochromatic light aa passes through the set of glass plates in the way shown in the figure. Just as in the ordinary transmission grating, the phase difference in the light emerging through consecutive steps must be $N\lambda$, where N is an integer, in any direction in which an image of the slit is formed by light of wave-length λ; but in the echelon grating we are only concerned with directions which make quite small angles, θ, with that of the incident beam. Now the difference in phase between one step and the next is easily seen to be
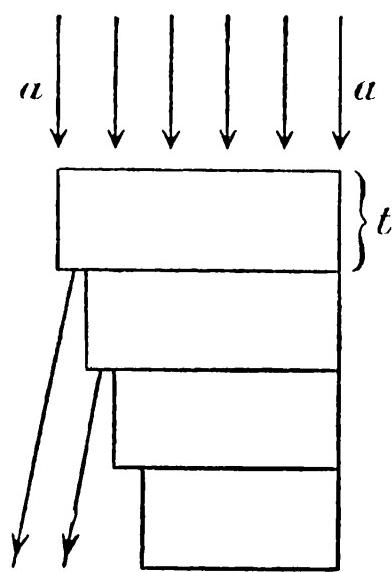
FIG. 31·1

$$\frac{(\mu - 1)t}{\cos \theta} + s\theta = N\lambda,$$

where t is the thickness of one of the plates, μ is the refractive index of the glass and s is the breadth of a step. Since θ is small, this may be written

$$(\mu - 1)t + s\theta = N\lambda. \quad . \quad . \quad . \quad (31\cdot1)$$

Hence

$$s\frac{d\theta}{d\lambda} = N - \frac{d\mu}{d\lambda}t.$$

If we multiply this by the number of steps, n, over which the beam spreads, we get

$$ns\frac{d\theta}{d\lambda} = n\left(N - \frac{d\mu}{d\lambda}t\right).$$

The left-hand member of this equation is the product of the breadth of the beam and the dispersion, and is therefore equal to the resolving power by (30·27). Hence

$$R = n\left(N - \frac{d\mu}{d\lambda}t\right). \quad . \quad . \quad . \quad (31\cdot11)$$

It is easy to see that the order N is enormous. In fact $N\lambda$ in (31·1) is of the order of magnitude of $(\mu - 1)t$, say 6 mm., and in the case of sodium light λ is roughly 6×10^{-5} cm. Therefore

N is about 10,000. With 30 plates therefore ($n = 30$) the resolving power would reach 300,000 without including the term $- nt \, d\mu/d\lambda$.

The resolving power of the echelon is so great that it can only be used to analyse light which is already highly homogeneous, and the echelon must consequently be placed on a spectroscope the resolving power of which is already great enough to admit through the echelon only a very narrow range of wave-lengths. When this condition is not satisfied serious overlapping occurs, with a consequent difficulty or impossibility of interpreting what is seen.

The echelon is very effective in showing up the fine structure of what, in apparatus of lower resolving power, appear as single lines.

If $d\theta_0$ be the angular separation between two consecutive orders N and $N + 1$, we have

$$(\mu - 1)t + s(\theta + d\theta_0) = (N + 1)\lambda,$$

and on subtracting (31·1) we get

$$s d\theta_0 = \lambda \quad . \quad . \quad . \quad . \quad . \quad \textbf{(31·12)}$$

and we realize that the next order on either side of $\theta = 0$ is absent, since by (30·24) the amplitude vanishes. On account of the relatively great width of s all the diffracted light is contained within so narrow an angle that in general only two orders of a line are seen, and therefore only a *single order* when $\theta = 0$. We can calculate the wave-length separation of two very close lines which appear in the same order in the following way. Differentiating (31·1) we get

$$d\mu.t + s d\theta = N d\lambda.$$

If now $d\theta = d\theta_0$ be the separation between consecutive orders, we may substitute for it the value λ/s (31·12). Thus

$$t.d\mu_0 + \lambda = N d\lambda_0, \quad . \quad . \quad . \quad . \quad \textbf{(31·13)}$$

where $d\mu_0$ and $d\lambda_0$ represent changes corresponding to the angular separation between the orders N and $N + 1$. Hence

$$t\frac{d\mu}{d\lambda}d\lambda_0 + \lambda = N d\lambda_0,$$

or

$$d\lambda_0 = \frac{\lambda}{N - t\dfrac{d\mu}{d\lambda}}.$$

Therefore

$$dλ_0 = \frac{λ^2}{Nλ - λt\dfrac{dμ}{dλ}}.$$

Now

$$Nλ = (μ - 1)t,$$

to a sufficiently close approximation by **(31·1)**. Therefore

$$dλ_0 = \frac{λ^2}{\left(μ - 1 - \dfrac{λdμ}{dt}\right)t}. \qquad . \quad . \quad . \quad \textbf{(31·14)}$$

and the wave-length difference corresponding to any other separation may be obtained by multiplying $dλ_0$ by $dθ/dθ_0$, where $dθ$ is the corresponding angular separation.

Recently W. E. Williams has succeeded in constructing echelon gratings of the reflecting type. This type of instrument is free from the disadvantage of the absorption in the plates of the transmission type of apparatus ; while its resolving power is appreciably greater (three or four times greater) than that of a transmission echelon with the same number of plates.

§ 31·2. Michelson's Interferometer

This instrument is of special interest on account of its application, in the year 1881, by Michelson himself, and later in collaboration with Morley, to investigate the relative motion of the earth and the aether. It is illustrated in Fig. 31·2. The source of light, S, is an extended bright surface, and we shall regard it as a plane surface perpendicular to the ray which is shown in the figure. It consists of two mirrors, A and B, the reflecting surfaces

Fig. 31·2

of which are perpendicular to one another and of two glass plates, O and C, of equal optical thickness, the parallel faces of which make the angle $π/4$ with the surfaces of the mirrors. One at least of the reflecting surfaces, A or B, can be moved in a direction perpendicular to its plane by means of an accurate screw. The beam of light from S travels as indicated
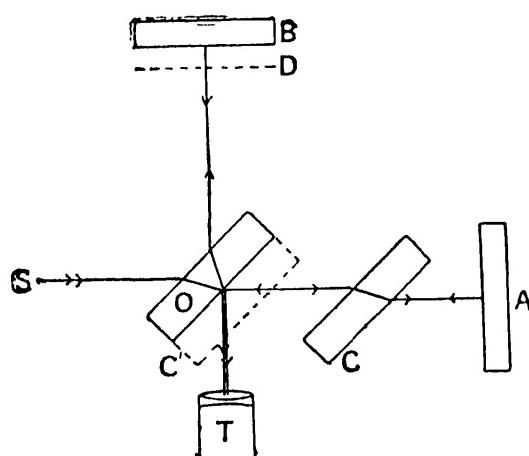
in the figure. It is partly reflected at the inner side of O, which is half silvered. The reflected beam travels to B, is there reflected and passes back through O to the observer. The portion of the beam from S which is transmitted through O passes on through C (which is called the compensating plate and is sometimes placed at C′ to insure parallelism with C), is reflected at A and passes back through C to be finally reflected at the half-silvered surface of O, so that it joins with the beam reflected from B. These two beams are coherent and interfere, producing characteristic fringes. We may regard them as coming from two images S′ and S″ (Fig. 31·21(a)), which, if the adjustment is perfect, will
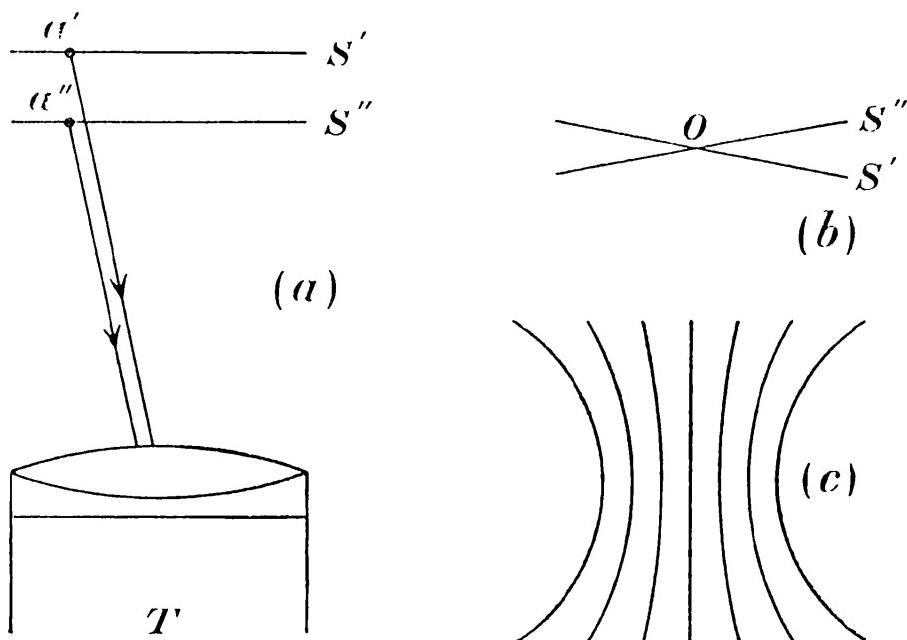
Fig. 31·21

have the form of plane and parallel surfaces. If a′ and a″ are the images of the same point a on the source, and if the light from it enters a telescope (adjusted for infinity) in the way illustrated in Fig. 31·21 (a), the two portions of it will reinforce one another when

$$(a'a'') \cos \theta = (n + \tfrac{1}{2})\lambda,^1 \quad . \quad . \quad . \quad \textbf{(31·2)}$$

n being an integer and θ the angle between the direction in which the light is travelling and the normal to B. The distance (a′a″) is obviously equal to 2t, where t is the perpendicular distance between the mirror B and the image, D, of the mirror A. Obviously a system of circular fringes will be seen by the observer. If the distance, t, were zero, the two surfaces would coincide under the ideal circumstances we are assuming, and the whole

---

¹ The ½ is due to a phase change of π, brought about by one of the reflexions as explained below.

field would appear dark. The reason for this is that the two cohering portions of the light from any point on S have travelled through the same optical distance ; but one of them suffers a change in phase, equal to π, on reflexion at O, while there is no such change in the case of the other (cf. **§ 25·5**). It is however impossible to adjust so that the two images are strictly parallel. They are, under the best circumstances, inclined at a small angle and intersect one another in an approximately straight line, O in Fig. 31·21 (*b*). When this line is in the field of view it can be recognized by using white light. It will appear as a dark line with a few coloured fringes on either side of it. Fringes can only be seen with white light when the difference of the optical paths of the interfering beams is very small. This is because each wave-length produces its own fringe system, and the separation of the fringes is the greater the greater the wave-length. The achromatic fringe, due as it is to the interference of beams which have travelled equal optical distances, marks the intersection of S' and S″, and is therefore practically straight. Hence the fringes in its neighbourhood are nearly straight. The appearance and curvature of fringes in this neighbourhood is illustrated in Fig. 31·21 (*c*).

The instrument is adjusted to produce fringes with monochromatic light such as that of a sodium flame by using a point source of light, i.e. by using a screen with a small hole in it. The hole in the screen admits light from a small portion (*a* in Fig. 31·21 (*a*)) of the bright surface. The inclination of one or other of the reflecting surfaces A or B is slightly altered, and the movable one is shifted by turning the screw, until *a'* and *a″* appear to coincide (absence of parallax).[1] If, when the screen is removed, fringes are not already visible, a very slight further change in the inclination of one of the mirrors will cause them to appear.

To make the adjustment for equal optical paths, monochromatic light is used to locate the short range of movement of the screw, which carries the movable mirror, through the region of approximately straight fringes from a point where their centres of curvature are on one side to a point where they are on the other side. White light is then turned on and the screw turned quite slowly through this range. The position of equal optical paths will be indicated by the sudden flashing out of the coloured fringes. The chief function of the compensating plate, C, is now explained. If it were absent, one of the optical paths would traverse the plate O twice, while the other would not

---

[1] Four images are seen ; but the coherent pair *a'* and *a″* can easily be picked out.

include any glass in its course. If therefore the two paths were equal for one wave-length, they would differ for others, and equality for all the wave-lengths, in the white light would be impossible. With the compensating plate each of the cohering beams travels the same distance through glass and through air and equality of optical path is achieved for all wave-lengths.

It is easy to apply the method of § 30·6 and show that the intensity varies through the fringe system according to a law like (30·63). The bright fringes are therefore rather broad and the direct resolving power of the interferometer is very small. Two lines so far separated in wave-length as the D lines of sodium, which are easily resolved by an ordinary prism spectroscope, give rise to systems of fringes which cannot be seen separately. They may overlap so well as to present the appearance of a single fringe system. If however the path difference is varied continuously the fringes gradually get out of coincidence, till one set of bright fringes falls over and fills up the dark spaces between those of the other, thus producing almost uniform illumination. Thus the **visibility** of the fringes changes progressively as the movable mirror travels. By studying these changes in visibility Michelson was able to disentangle the fine structure of many spectral lines. This variation of the visibility of the fringes endows the apparatus with an indirect resolving power of great magnitude—it is only limited by the precision with which the visibility can be estimated numerically.

§ 31·3. Velocity of the Aether Relative to the Earth

The old view that light waves, or electromagnetic waves, are propagated by a luminiferous medium (aether), raised the question as to whether the earth in its motion carries along with it the aether which is round about it ; or, if not, what is the relative velocity of the two. Michelson attempted to answer these questions by searching for small variations in the velocity of light, relative to the earth, in different directions. The interferometer just described was used for this purpose. It was mounted on a firm base which floated on mercury. It was thus possible to turn it so that its arms pointed in various directions without introducing any shift of the fringes due to strain. Any dependence of the velocity of light on the direction would show itself by a shift of the fringes. Such a movement of the fringes was not observed, and Michelson concluded that the aether was carried along with the earth. This explanation, however, raised difficulties in another direction, namely in connexion with the phenomenon of the aberration of light. Michelson's negative

result will be discussed in § 33·4 and in the description of the theory of relativity.

§ 31·4. Measurement of Length by Counting Light Waves

To count the number of waves of a homogeneous source of light (the light of the cadmium red line for example) Michelson replaced the single mirror A (Fig. 31·2) by two **etalons** placed side by side. Each etalon consisted of two mirrors, both of them parallel to the position of the mirror A (Fig. 31·2) which has been replaced; but one of them above the other as illustrated by A and B in Fig. 31·4. The two etalons were placed side by side. In the first instance the surfaces of the lower mirrors A and A′ were brought into coincidence by using white light fringes. The perpendicular distance AB of one etalon was made as nearly as practicable double the distance A′B′ between the mirrors of the second etalon, so that in the first setting the mirror B′ would be in the position indicated in Fig. 31·4. Both the mirror B (Fig. 31·2) and the two etalons can be moved by accurate screws. All these screws were arranged, by the introduction of an extra reflecting surface, to travel parallel to one another. Having brought the surfaces A and A′ (Fig. 31·4) into coincidence, the mirror B (Fig. 31·2) was moved back until it formed white light fringes with B′ (Fig. 31·4). The shorter etalon was then moved till A′ reached the position which had been occupied by B′, when the latter reflecting surface would be separated by a minute distance from the reflecting surface B. The number of wave-lengths contained in this small distance was determined by counting fringes with the cadmium light, each terminal position being determined, of course, by the observation of fringes with white light. If now the number of waves in the shorter etalon A′B′ has been previously determined, those in the longer one AB can be obtained, since the latter number will be twice the former, plus or minus the number which has been determined by direct counting. Michelson used etalons the lengths of which were 10 cm., 5 cm., 2·5 cm. and so on ; the shortest being 0·39 mm. long. The number of waves in the last named was found by direct counting. This enabled the number of waves in the 0·781 mm. etalon to be found, and

Fig. 31·4

so onwards till the number in the 10 cm. one had been found. Last of all this etalon was compared with the prototype metre by means of a comparator. For fuller details the work on interferometry by W. E. Williams may be consulted.

§ 31·5. THE INTERFEROMETER OF FABRY AND PEROT— LUMMER PLATE

The interferometer of Fabry and Perot consists of two parallel glass plates, half silvered on their inner faces. Sometimes the distance between them can be varied by means of an accurate screw, and sometimes it is fixed (Fabry and Perot etalon). A beam of light passed through the pair of plates undergoes many reflexions at the half-silvered surfaces and the coherent multiple beams thus produced give rise to a system of interference fringes. We may regard the Fabry and Perot interferometer, and also the Lummer plate, as of the nature of an echelon grating with a vaguely defined number of steps—since each pair of internal reflexions brings about a phase change corresponding to that between the consecute steps of a transmission echelon; but the progressive loss of light at each reflexion introduces a feature which might be simulated by an echelon with a very large number of imperfectly transparent plates.

The bright fringes produced by these interferometers are very narrow, compared with the separation between fringes of consecutive orders. It is easy to find the reason for this without going deeply into theoretical details. Any portion of the incident beam is immensely broadened by the multiple reflexions and by **(30·24)** the angular breadth of a fringe is effectively determined by λ/breadth.

The Lummer plate closely resembles the Fabry and Perot instrument. In it the multiple reflexions are internal reflexions at the parallel surfaces of a glass plate. It is therefore not possible to alter the perpendicular distance between the reflecting surfaces. It has the advantage that the angle of reflexion in the glass can be adjusted to approach very closely to the critical angle, and very strong reflexion can thus be attained; but on the other hand it is very difficult to get a piece of glass with the optical homogeneity which is essential, and there is also the disadvantage of the absorption in the glass.

It is easy to find a suitable expression for the separation $d\lambda_0$ between two adjacent orders N and $N + 1$.

Let μ be the refractive index of the medium between the reflecting surfaces of the Fabry and Perot interferometer (or of the glass of the Lummer plate). If the angle between the

direction of the beam *in the medium* and the normal be θ, the path difference is easily seen to be $2\mu t \cos \theta$. Therefore the Nth order fringe must conform to the equation

$$N\lambda = 2\mu t \cos \theta, \quad . \quad . \quad . \quad . \quad . \quad \textbf{(31·5)}$$

where t is the thickness of the medium, and consequently

$$\lambda = -2\mu t \sin \theta \, d\theta_0, \quad . \quad . \quad . \quad \textbf{(31·51)}$$

where $d\theta_0$ is the angular separation of fringes of consecutive orders. The difference, $d\lambda_0$, in wave-length, corresponding to $d\theta_0$ is obtained by differentiating **(31·5)**

$$N d\lambda_0 = 2 d\mu_0 t \cos \theta - 2\mu t \sin \theta \, d\theta_0, \quad . \quad . \quad \textbf{(31·52)}$$

where $d\mu_0$ is the change of refractive index corresponding to $d\lambda_0$. We may write **(31·52)** in the form :

$$N d\lambda_0 = 2\frac{d\mu}{d\lambda} d\lambda_0 t \cos \theta + \lambda,$$

by **(31·51)**. Therefore

$$d\lambda_0 = \frac{\lambda}{N - 2\dfrac{d\mu}{d\lambda} t \cos \theta},$$

$$d\lambda_0 = \frac{\lambda^2}{N\lambda - 2\dfrac{d\mu}{d\lambda}\lambda t \cos \theta},$$

and finally, by **(31·5)**

$$d\lambda_0 = \frac{\lambda^2}{2t \cos \theta \left(\mu - \lambda\dfrac{d\mu}{d\lambda} \right)}. \quad . \quad . \quad \textbf{(31·52)}$$

In the case of the Lummer plate θ is very near the critical angle of the glass ; but with the Fabry and Perot instrument it is nearly zero, and $\lambda d\mu/d\lambda$ is negligible by comparison with μ, which is approximately unity, so that **(31·52)** becomes

$$d\lambda_0 = \frac{\lambda^2}{2t}, \quad . \quad . \quad . \quad . \quad \textbf{(31·53)}$$

to a sufficiently close approximation.

§ 31·6.   Brewster's Fringes

Imagine two Fabry-Perot interferometers, or etalons, to be placed in series, i.e. to be placed one behind the other, so that a beam of light can pass in succession through both of them. Let us further suppose that one has exactly twice the thickness of the other. Under these circumstances coloured fringes are

produced when white light is used. These are known as **Brewster's fringes.** The way in which they are produced is illustrated in Fig. 31·6. A portion of the light coming from some point on the source may be reflected at A and B in the first etalon, and then pass straight on through the second one, while another portion of it may pass through the first etalon and be reflected in the second one at the points D, E, F and G; so that the two paths ABC and $DEFGH$ are equal.

An alternative method of finding the number of wave-lengths of very homogeneous light waves (such as those of cadmium red light) in the prototype metre was used by Fabry, Perot and Benoit. A number of etalons were placed in series. In order of length these were 6·25, 12·5, 25, 50 and 100 cms. They might have been made so that each was *exactly* double the length of its immediate predecessor, by using Brewster's fringes. To avoid the great trouble of doing this directly, thin half-silvered air wedges were added to one or the other by suitably placed mirrors. The number of waves in the shortest etalon is equal to an integer plus some fraction. This integer can be correctly found from the known value of the wave-length of the light under investigation, and the fraction can be found by measuring the angular diameters of the circular fringes. Any possibility of error may be eliminated by calculating the fractional parts to be expected with other known wave-lengths, and testing whether they are in agreement with observation. Now knowing the number of waves in the shortest etalon, we get the number of the next one by multiplying by two and adding or subtracting, as the case may be, the number in the equalizing air wedge. Finally the number contained in the longest etalon (as near as may be one metre in length) is found, and by comparing this with the prototype metre the number of waves in that may be determined.
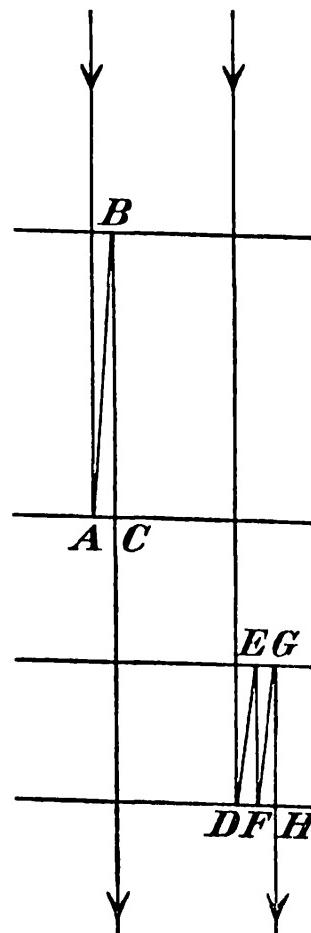
Fig. 31·6

BIBLIOGRAPHY

Schuster and Nicholson : An Introduction to the Theory of Optics. (Arnold, 1924.)

W. E. Williams : Applications of Interferometry. (Methuen.)

R. W. Wood : Physical Optics. (The Macmillan Co.)

P. Drude : Lehrbuch.

HUYGENS' PRINCIPLE

§ 31·7. ORIGINAL FORM OF HUYGENS' PRINCIPLE—FRESNEL'S IMPROVEMENT

THE explanation of the rectilinear propagation of light in a homogeneous medium was the main obstacle in the way of the general adoption of the undulatory theory. To account for it Huygens enunciated the principle named after him. In its original form it asserts that each point on a wave front may be regarded as the source of a secondary wave, and that only the surface which envelops the secondary wave surfaces can produce appreciable effects. So stated the principle does in fact lead to rectilinear propagation. It also leads to the correct laws of reflexion and refraction, and to the correct relationship between refractive indices and phase velocities. The principle was improved by Fresnel, who appreciated that the interference of the elementary or secondary waves with one another not only renders Huygens' assumption about the enveloping surface unnecessary ; but explains the diffraction phenomena which are observed when a beam of light is limited by the boundary of an aperture or of an opaque obstacle, and the observed fact that rectilinear propagation is only a limit approached as the wave-length gets smaller and smaller by comparison with the dimensions of the lenses, apertures, or the region under examination.

Huygens' principle in its original form leads to a propagation backwards towards the source, and Fresnel's improvement does not get rid of this weakness. Fresnel's theory also introduces an error of $\pi/4$ in the phase of the luminous vibrations.

§ 31·8. RIGOROUS FORM OF HUYGENS' PRINCIPLE

Huygens' principle expresses, or attempts to express, the luminous disturbance at a given point (O in Fig. 31·8) as a sum of contributions from the elements of a surface, S, which we

may suppose to surround O. If there be such an expression, it must be obtainable from the wave equation :

$$\frac{\partial^2 \mathcal{E}_x}{\partial t^2} = v^2 \nabla^2 \mathcal{E}_x, \qquad \qquad \textbf{(31·8)}$$

and the two associated equations, in which $\mathcal{E}_x$, $\mathcal{E}_y$ and $\mathcal{E}_z$ are the components of the electric field intensity (or of the electric displacement) and v is the phase velocity. We are assuming, of course, a homogeneous and isotropic medium. We can, in fact, express the solution of **(31·8)** in the form of a surface integral by means of **(27·55)**. In the latter formula there is a volume integral due to the electric density, ρ. This integral is absent in the problem before us, owing to the absence of the corresponding term in the differential equation **(31·8)**. We shall use rectangular co-ordinates and place the origin at O. We then have, by **(27·55)**, for

Fig. 31·8

$\mathcal{E}_{0,0,0,t}$, i.e. for the electric intensity at the origin, O, at the instant, t :

$$\mathcal{E}_{0;0,0;t} = \frac{1}{4\pi} \int\int \left\{ \frac{1}{r}\left[\frac{\partial \mathcal{E}}{\partial n}\right] - [\mathcal{E}]\frac{\partial\left(\frac{1}{r}\right)}{\partial n} + \frac{1}{vr}\left[\frac{\partial \mathcal{E}}{\partial t}\right]\frac{\partial r}{\partial n} \right\} dS, \quad \textbf{(31·81)}$$

in which dS is an element of the surface, S ; r is the radius vector from O to the element, dS, and n is the outward normal. The integration is extended over the surface, S. If now the point source at σ consists of harmonic vibrations with the period, τ, the electric intensity at a point, p, on the surface can be expressed by

$$\mathcal{E} = \frac{a}{r_1} \cos 2\pi\left(\frac{t}{\tau} - \frac{r_1}{\lambda}\right), \qquad \qquad \textbf{(31·82)}$$

where a is the amplitude at the unit distance from σ and r_1 is the distance from σ to p (cf. **§ 9·1**). Now

$$\frac{\partial \mathcal{E}}{\partial n} = \frac{\partial \mathcal{E}}{\partial r_1}\frac{\partial r_1}{\partial n},$$

therefore

$$\frac{\partial \mathcal{E}}{\partial n} = \left\{ -\frac{a}{r_1^2}\cos 2\pi\left(\frac{t}{\tau} - \frac{r_1}{\lambda}\right) + \frac{2\pi a}{\lambda r_1}\sin 2\pi\left(\frac{t}{\tau} - \frac{r_1}{\lambda}\right) \right\}\frac{\partial r_1}{\partial n}.$$
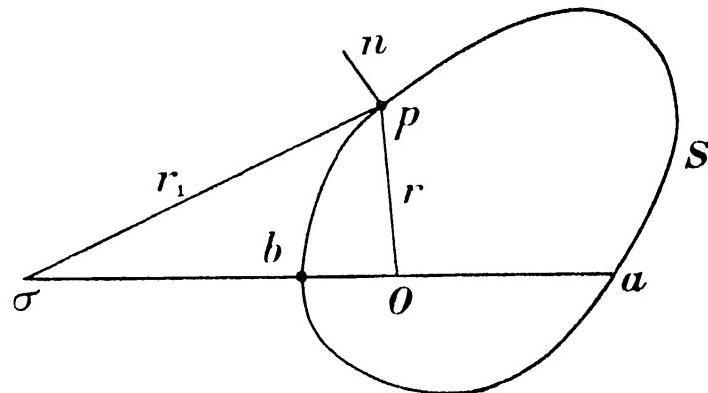
We shall assume λ to be very small by comparison with r_1 and remember that $\partial r_1/\partial n = \cos (n, r_1)$. Therefore

$$\frac{\partial \mathcal{E}}{\partial n} = \frac{2\pi a}{\lambda r_1} \cos (n, r_1) \sin 2\pi \left(\frac{t}{\tau} - \frac{r_1}{\lambda}\right).$$

Now the significance of the brackets [ ] in (31·81) is that t is replaced by $t - r/v$. Therefore

$$\left[\frac{\partial \mathcal{E}}{\partial n}\right] = \frac{2\pi a}{\lambda r_1} \cos (n, r_1) \sin 2\pi \left(\frac{t}{\tau} - \frac{r_1 + r}{\lambda}\right). \qquad . \quad (31\cdot83)$$

Similarly,

$$\frac{1}{vr}\left[\frac{\partial \mathcal{E}}{\partial t}\right]\frac{\partial r}{\partial n} = - \frac{2\pi a}{\lambda r_1 r} \cos (n, r) \sin 2\pi \left(\frac{t}{\tau} - \frac{r_1 + r}{\lambda}\right). \quad (31\cdot84)$$

The middle term in (31·81) can be neglected when $\lambda \ll r_1$, and therefore we obtain, on substituting (31·83) and (31·84) in (31·81) :

$$\boxed{\mathcal{E}_{0,\,0,\,0,\,t} = \frac{a}{2\lambda}\iint \frac{(\cos (n,\, r_1) - \cos (n, r)\,)}{r_1 r} \sin 2\pi \left(\frac{t}{\tau} - \frac{r_1 + r}{\lambda}\right) dS.} \quad (31\cdot85)$$

This is Huygens' principle in the rigorous form first given by Kirchhoff.[1] It is free from the defects in the earlier forms. There is, for example, no backward propagation to O from the element at a (Fig. 31·8), since for this element $\cos (n, r_1) = \cos (n, r)$. The phase at O is also correctly given, since by (31·82) we must have at O

$$\mathcal{E} = \frac{a}{r_1 + r} \cos 2\pi \left(\frac{t}{\tau} - \frac{r_1 + r}{\lambda}\right),$$

while the integration in (31·85) transforms

$$\sin 2\pi(t/\tau - (r_1 + r)/\lambda)$$
into
$$\cos 2\pi(t/\tau - (r_1 + r)/\lambda).$$

§ 31·9. Application to a Rectangular Aperture

Imagine an infinitely extended plane screen with a rectangular aperture in it of length and breadth A and B respectively. Further, suppose a point source, like that in the last section, situated on the straight line or axis, through the middle point of the aperture and perpendicular to the plane of the screen. We shall study the case where the point source is very distant from the aperture, and inquire about the amplitude and in-

[1] G. Kirchhoff: *Vorlesungen über Mathematische Optik.*

tensity of the light at a very distant point on the straight line which passes through the centre of the aperture, and makes the angle θ with the axial direction. We may therefore regard the incident beam as a parallel beam. The phenomena of the diffraction of parallel beams are sometimes termed **Fraunhofer diffraction phenomena,** while other diffraction phenomena are termed **Fresnel diffraction phenomena.** We shall not deal at all with the latter, but refer the reader to Drude's *Lehrbuch* and to the other works mentioned at the end of this chapter.

In the problem before us, the integration of (31·85) is extended over the area of the rectangular aperture and is simplified to

$$\mathcal{E} = -\frac{a}{2\lambda}(1 + \cos\theta)\int\int \sin 2\pi\left(\frac{t}{\tau} - \frac{r_1 + r}{\lambda}\right)dS, \quad . \quad (31 \cdot 9)$$

since $\cos(n, r_1) = -1$, and $\cos(n, r) = \cos\theta$. Our task is to find the amplitude and intensity, i.e. $\mathcal{E}$ and $\mathcal{E}^2$, when r and r_1 are very great. We may write (31·9) in the simpler form :

$$\mathcal{E} = a'\int\int \sin 2\pi\left(\frac{t}{\tau} - \frac{r_1 + r}{\lambda}\right)dS, \quad . \quad . \quad (31 \cdot 91)$$

where $\qquad a' = -a(1 + \cos\theta)/2\lambda. \quad . \quad . \quad . \quad$ (31.911)

It is convenient to replace r by $r_0 + r$, where r_0 means the distance from the *centre* of the aperture to the point where we are calculating the intensity. Equation (31·91) may therefore be written :

$$\mathcal{E} = a'\sin 2\pi\left(\frac{t}{\tau} - \frac{r_1 + r_0}{\lambda}\right)\int\int \cos\frac{2\pi r}{\lambda}dS$$

$$- a'\cos 2\pi\left(\frac{t}{\tau} - \frac{r_1 + r_0}{\lambda}\right)\int\int \sin\frac{2\pi r}{\lambda}dS.$$

If we write

$$\left.\begin{array}{l} C = \int\int \cos\dfrac{2\pi r}{\lambda} \, dS, \\[2mm] D = \int\int \sin\dfrac{2\pi r}{\lambda} \, dS, \end{array}\right\} \quad . \quad . \quad . \quad . \quad (31 \cdot 92)$$

we have for $\mathcal{E}$:

$$\mathcal{E} = a'C\sin 2\pi\left(\frac{t}{\tau} - \frac{r_1 + r_0}{\lambda}\right) - a'D\cos 2\pi\left(\frac{t}{\tau} - \frac{r_1 + r_0}{\lambda}\right).$$

Therefore

$$\mathcal{E} = \sqrt{a'^2 C^2 + a'^2 D^2}\,\bigg|\,\sin 2\pi\left(\frac{t}{\tau} - \frac{r_1 + r_0}{\lambda} - \phi\right), \quad . \quad (31 \cdot 93)$$

where $\tan\phi = D/C$.

We shall use the letter x for distances measured across the aperture parallel to the length A, and the letter y for distances measured parallel to the breadth B. In both cases we shall measure from the centre of the aperture, so that x varies from $-A/2$ to $+A/2$, while y varies from $-B/2$ to $+B/2$. In fact we may imagine rectangular co-ordinates with the origin in the centre of the aperture and the X and Y axes parallel to its length and breadth respectively. Now if ρ be the distance of any point in the plane of the aperture from the central point, the corresponding value of r (in **31·92**) is clearly

$$r = \rho \sin \theta,$$

while

$$\rho = x \cos \varepsilon + y \cos \eta, \qquad . \quad . \quad . \quad \textbf{(31·94)}$$

where ε and η are the angles between the direction of ρ and of the X and Y axes respectively. Obviously then

$$r = \mu x + \nu y,$$

if

$$\mu = \sin \theta \cos \varepsilon,$$

and

$$\nu = \sin \theta \cos \eta. \qquad . \quad . \quad . \quad \textbf{(31·95)}$$

Finally we may replace dS by $dx\, dy$. Substituting in (**31·92**) we get

$$C = \int\!\!\int \cos \frac{2\pi}{\lambda}(\mu x + \nu y) dx\, dy$$

$$D = \int\!\!\int \sin \frac{2\pi}{\lambda}(\mu x + \nu y) dx\, dy \qquad . \quad . \quad . \quad \textbf{(31·96)}$$

Thus, on expanding the *cosine* and *sine*, we get

$$C = \int\!\!\int \cos \frac{2\pi\mu x}{\lambda} \cos \frac{2\pi\nu y}{\lambda}\, dx\, dy,$$

$$D = \qquad 0, \qquad . \quad . \quad \textbf{(31·961)}$$

since the integrals containing the *sine* of $2\pi\mu x/\lambda$, or of $2\pi\nu y/\lambda$, vanish. We easily find for C :

$$C = \left\{ \frac{\sin \dfrac{\pi\mu A}{\lambda}}{\dfrac{\pi\mu}{\lambda}} \right\} \left\{ \frac{\sin \dfrac{\pi\nu B}{\lambda}}{\dfrac{\pi\nu}{\lambda}} \right\},$$

and the required amplitude, namely $\sqrt{a'^2 C^2 + a'^2 D^2}\,|$, is

$$ABa' \left\{ \frac{\sin \dfrac{\pi\mu A}{\lambda}}{\dfrac{\pi\mu A}{\lambda}} \right\} \left\{ \frac{\sin \dfrac{\pi\nu B}{\lambda}}{\dfrac{\pi\nu B}{\lambda}} \right\}, \quad . \quad . \quad . \quad \textbf{(31·97)}$$

in which a' is numerically equal to a' in (31·91). When A is very great the aperture becomes a slit and we meet again the case of § 30·2. The factor $\sin(\pi\mu A/\lambda)/(\pi\mu A/\lambda)$ can now only differ from zero if μA approaches zero as A becomes very great, in which case it has the limiting value unity. Consequently $\mu = 0$ and therefore, when $\sin\theta$ differs from zero, $\cos\varepsilon = 0$ and $\cos\eta = 1$, by (31·94). In these circumstances $\rho = y$, $\nu = \sin\theta$ and $r = y\sin\theta$. Thus the amplitude becomes

$$ABa' \frac{\sin\left(\dfrac{\pi\sin\theta B}{\lambda}\right)}{\dfrac{\pi\sin\theta B}{\lambda}}. \quad . \quad . \quad . \quad . \quad \textbf{(31·98)}$$

and the first dark fringe appears where

$$\pi\sin\theta . B/\lambda = \pi,$$

or

$$\sin\theta = \lambda/B,$$

in agreement with what we found in § 30·2. Moreover, according to (30·21) the resulting phase is that of the light which passes through the central part of the slit. We find the same thing again here, since $\tan\phi = D/C = 0$, on account of $D = 0$, and we may replace ϕ in (31·93) by zero. The present theory is, however, more complete than that of § 30·2, because we have found an expression for the factor $a'AB$ in the amplitude which is valid for any angle, θ. Equation (31·911) shows that it is sensibly constant for small values of θ, as we assumed in § 30·2. In the special case where $\theta = 0$, r in (31·92) is zero, and consequently $C = AB$ and the amplitude becomes $a'AB$.

§ 32. Babinet's Principle

When the wave-length in a beam of monochromatic light which passes through an aperture, A, in an infinitely extended screen, is exceedingly short by comparison with the dimensions of the aperture, the illuminated surface of a second screen on which the light falls after passing the aperture will be limited by an almost sharp boundary, beyond which will be almost complete darkness. The boundary will be all the more sharply defined, and the darkness the more complete, the shorter the wave-length by comparison with the dimensions of A. This can be proved at once for a rectangular aperture by observing that (31·97) approaches zero when λ approaches zero *except when θ is zero*, and for an aperture of any shape by the use of

equation (31·9) (cf. § 28·2). If, however, portions of the area, A, be occupied by opaque screens, diffraction effects may be produced in the previously dark region, since the wave-length may now no longer be very short by comparison with the dimensions of the screens or of the resulting apertures. Babinet's principle asserts that these diffraction effects will be identical in all respects with those which result when the parts of A, previously occupied by opaque screens, are made transparent and the remaining, previously transparent, parts made opaque. This may be proved in the following way : In the first case the intensity at any point *in the region of the geometrical shadow* is proportional to

$$C_1{}^2 + D_1{}^2,$$

where C_1 and D_1 have the meanings explained in § 31·9. Similarly, with the complementary screens the intensity at such a point is proportional to

$$C_2{}^2 + D_2{}^2.$$

Now with the full illumination, i.e. with the whole aperture, A, open, the intensity at the same point will be proportional to

$$(C_1 + C_2)^2 + (D_1 + D_2)^2,$$

and this must be equal to zero, as we have seen. Being made up of two positive terms, the expression can only vanish if each term separately vanishes. Therefore

$$C_1 = - C_2,$$
$$D_1 = - D_2,$$

and thus

$$C_1{}^2 + D_1{}^2 = C_2{}^2 + D_2{}^2,$$

and the principle is established. It is important to notice that its validity is limited to cases where, with full illumination, i.e. with the complementary apertures both open, there is complete, or practically complete darkness.

BIBLIOGRAPHY

C. Huygens : Traité de la Lumière, 1690. This work has been translated by Silvanus Thompson under the title *Treatise on Light*. (Macmillan.)

G. Kirchhoff : Vorlesungen über Mathematische Optik. (Teubner, Leipzig.)

P. Drude : Lehrbuch. This work contains an admirable account of Huygens' principle and its applications.

A. Schuster : An Introduction to the Theory of Optics.

CHAPTER XIV

PROPAGATION OF LIGHT IN CRYSTALLINE MEDIA—POLARIZED LIGHT

§ 32·1. Phase Velocity of Plane Waves

WE learned in § 25·4 that the phase velocity of a plane wave in an anisotropic medium is governed by the equation

$$\frac{l^2}{v^2 - a^2} + \frac{m^2}{v^2 - b^2} + \frac{n^2}{v^2 - c^2} = 0,$$

in which l, m and n are the direction cosines specifying the direction of propagation, and a, b and c are the principal velocities defined by (25·441) and (25·442). If the equation be written in the form :

$$l^2(v^2 - b^2)(v^2 - c^2) + m^2(v^2 - a^2)(v^2 - c^2)$$
$$+ n^2(v^2 - a^2)(v^2 - b^2) = 0, \quad . \quad . \quad . \quad \textbf{(32·1)}$$

it will be seen to be quadratic in v^2, so that there are two values of v^2 for a given direction (l, m, n). For each value of v^2 we have, of course, two numerically equal values of v, a positive and a negative one. The significance of this is simply that a plane wave travels with the same velocity, whether travelling in the direction (l, m, n) or in the opposite one, and we shall ignore the negative value in what follows. Let us first investigate velocities in directions perpendicular to the X principal axis, i.e. in directions for which $l = 0$. Equation (32·1) now becomes

$$(v^2 - a^2)(v^2 - m^2c^2 - n^2b^2) = 0, \quad . \quad . \quad \textbf{(32·11)}$$

since $m^2 + n^2 = 1$, and we infer that a plane wave is propagated in every one of these directions with the velocity a. We have met with this plane wave already in § 25·4. It is associated with vibrations (displacements) parallel to the X principal axis. The remaining factor in (32·11) gives the velocity

$$v = \sqrt{m^2c^2 + n^2b^2 \mid} = \sqrt{c^2 + n^2(b^2 - c^2) \mid} ; \quad . \quad \textbf{(32·12)}$$

19   289

so that in directions perpendicular to the X principal axis plane waves are propagated with velocities which range from c to b according to the value of n. When $n = 0$ (Y direction), $v = c$, and we have a plane wave in which the displacements are parallel to the Z principal axis ; when $n = 1$ (Z direction), $v = b$ and the direction of displacement is parallel to the Y principal axis. The direction of displacement in the general case will be investigated in § 32·6. We may say then, in general, that a plane wave, on entering a crystalline medium, splits up into two plane waves which travel with different velocities, each of which is associated with a definite direction of displacement.

We may deal similarly with the groups of directions associated with $m = 0$ and $n = 0$. The former calls for special attention because the principal velocity, b, of plane waves, the direction of displacement in which is parallel to the Y principal axis, is intermediate in value between a and c. The two velocities in a direction perpendicular to the Y axis are

$$v = b \text{ and } v = \sqrt{n^2 a^2 + l^2 c^2 \,|} \,, \qquad . \quad . \quad \textbf{(32·13)}$$

and therefore there are directions for which the two values of v are identical, namely

$$\left. \begin{aligned} l_0 &= \pm \sqrt{\left|\frac{a^2 - b^2}{a^2 - c^2}\right|} \,, \\ m_0 &= 0, \\ n_0 &= \pm \sqrt{\left|\frac{b^2 - c^2}{a^2 - c^2}\right|} \,. \end{aligned} \right\} \qquad . \quad . \quad . \quad \textbf{(32·14)}$$

The two lines or directions thus determined are called the **optic axes** of the medium. Obviously when $a^2 = b^2$, or when $b^2 = c^2$, they coincide ; as is otherwise evident when we make $a = b$, for example, in **(32·1)**, which then becomes

$$(v^2 - a^2)\{v^2 - c^2 + n^2(c^2 - a^2)\} = 0, \quad . \quad . \quad \textbf{(32·15)}$$

and it is clear that both velocities are equal to a when $n = 1$. Iceland spar and quartz are examples of **uniaxial crystals,** or crystals in which the two axes coincide. In the former $b = c$, i.e. the two smaller principal velocities are equal, while in quartz $a = b$.

§ 32·2. The Wave Surface—Rays in Crystalline Media

When monochromatic light waves are generated at a point, O (Fig. 32·2), a wave crest created there travels outwards in all directions and the points reached by it in the unit time constitute

a locus called the **wave surface.** It is obviously a spherical surface in the case of homogeneous isotropic media, and there is an alternative way of regarding it, whether we are dealing with isotropic or anisotropic media. Any small part of it may be thought of as part of a plane wave which touches the wave surface at the point A, where the surface and the plane coincide. The length of the perpendicular from O to the plane is equal to the wave velocity, v, in the direction of the perpendicular. Conse-

FIG. 32·2

quently the wave surface is the surface which envelopes or is touched by all the planes

$$xl + ym + zn = v, \quad . \quad . \quad . \quad . \quad \textbf{(32·2)}$$

each plane being determined by the direction cosines (l, m, n) and the length, v, of the perpendicular from the origin. We shall call the *radii vectores* **rays** and the length, r, of each one, the **ray velocity.** It may be pointed out that the rays are rays in the sense of **§ 28·2** when the conditions appropriate to geometrical optics are satisfied. The angle, ϕ, between the direction of the ray and that of the propagation of the plane wave is determined by the equation

$$\cos \phi = v/r. \quad . \quad . \quad . \quad . \quad \textbf{(32·21)}$$

§ 32·3. Equation of the Wave Surface

Since the wave surface is the surface touched by all the planes **(32·2)** we may regard any point (x, y, z) on it as the point of intersection of **(32·2)** and all its neighbouring planes the equation of any one of which may be written :

$$x(l + \delta l) + y(m + \delta m) + z(n + \delta n) = v + \delta v.$$

Consequently for a given point (x, y, z) on the wave surface

$$x\delta l + y\delta m + z\delta n = \delta v. . \quad . \quad . \quad . \quad \textbf{(32·3)}$$

Now the variations δl, δm, δn and δv cannot all be chosen arbitrarily, on account of the relationships

$$l^2 + m^2 + n^2 = 1,$$

$$\frac{l^2}{v^2 - a^2} + \frac{m^2}{v^2 - b^2} + \frac{n^2}{v^2 - c^2} = 0.$$
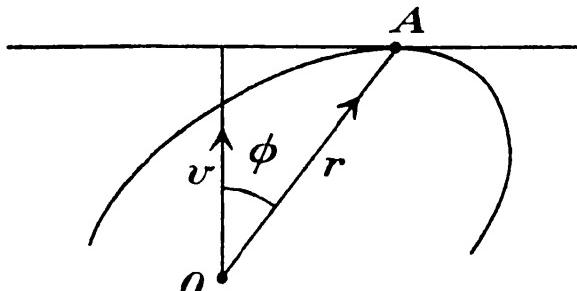
They are consequently subject to the conditions :

$$l\delta l + m\delta m + n\delta n = 0,$$

$$\left.\frac{l\delta l}{v^2 - \mathbf{a}^2} + \frac{m\delta m}{v^2 - \mathbf{b}^2} + \frac{n\delta n}{v^2 - \mathbf{c}^2} = Kv\delta v,\right\} \qquad \textbf{(32·301)}$$

where

$$K \equiv \frac{l^2}{(v^2 - \mathbf{a}^2)^2} + \frac{m^2}{(v^2 - \mathbf{b}^2)^2} + \frac{n^2}{(v^2 - \mathbf{c}^2)^2}. \qquad . \quad \textbf{(32·31)}$$

If we multiply **(32·3)** by $- Kv$ and the first of the equations **(32·301)** by an undetermined factor, A, and add them to the second equation **(32·301)**, we get

$$\left(Al - Kvx + \frac{l}{v^2 - \mathbf{a}^2}\right)\delta l + \left(\begin{array}{c}\text{terms containing}\\ \delta m \text{ and } \delta n\end{array}\right) = 0.$$

We get rid of the term containing δl by choosing A, so that

$$Al - Kvx + \frac{l}{v^2 - \mathbf{a}^2} = 0, \quad . \quad . \quad . \quad \textbf{(32·32)}$$

and since δm and δn in the two remaining terms are arbitrary, it follows that

$$\left.\begin{array}{c}Am - Kvy + \dfrac{m}{v^2 - \mathbf{b}^2} = 0,\\[2mm] An - Kvz + \dfrac{n}{v^2 - \mathbf{c}^2} = 0.\end{array}\right\} \quad . \quad . \quad \textbf{(32·321)}$$

If we now multiply equations **(32·32)** and **(32·321)** by l, m and n respectively and add, we find

$$A - Kv^2 = 0, \quad . \quad . \quad . \quad . \quad \textbf{(32·33)}$$

in consequence of **(32·2)**. On substituting this value of A in **(32·32)** and **(32·321)** we get

$$\left.\begin{array}{l}Kv(x - vl) \ = \dfrac{l}{v^2 - \mathbf{a}^2},\\[3mm] Kv(y - vm) = \dfrac{m}{v^2 - \mathbf{b}^2},\\[3mm] Kv(z - vn) \ = \dfrac{n}{v^2 - \mathbf{c}^2},\end{array}\right\} \quad . \quad . \quad . \quad \textbf{(32·34)}$$

and on squaring these and adding :

$$Kv^2(r^2 - v^2) = 1, \quad . \quad . \quad . \quad . \quad \textbf{(32·35)}$$

in which $r^2 = x^2 + y^2 + z^2$. If we now substitute the value of K given in **(32·35)** in equations **(32·34)** we obtain

$$\frac{x - vl}{r^2 - v^2} = \frac{vl}{v^2 - \mathbf{a}^2}$$

and two similar equations. Therefore

$$\left.\begin{aligned}
\frac{x - vl}{r^2 - v^2} &= \frac{vl}{v^2 - \mathbf{a}^2} = \frac{x}{r^2 - \mathbf{a}^2}, \\[2mm]
\frac{y - vm}{r^2 - v^2} &= \frac{vm}{v^2 - \mathbf{b}^2} = \frac{y}{r^2 - \mathbf{b}^2}, \\[2mm]
\frac{z - vn}{r^2 - v^2} &= \frac{vn}{v^2 - \mathbf{c}^2} = \frac{z}{r^2 - \mathbf{c}^2}.
\end{aligned}\right\} \qquad . \quad . \quad . \quad \textbf{(32·36)}$$

On multiplying by x, y and z respectively and adding, we find

$$1 = \frac{x^2}{r^2 - \mathbf{a}^2} + \frac{y^2}{r^2 - \mathbf{b}^2} + \frac{z^2}{r^2 - \mathbf{c}^2}, \quad . \quad . \quad \textbf{(32·37)}$$

which is the equation of the wave surface. If we replace the left-hand side by $(x^2 + y^2 + z^2)/r^2$, we see that this equation may be written in the form :

$$\frac{\mathbf{a}^2 x^2}{r^2 - \mathbf{a}^2} + \frac{\mathbf{b}^2 y^2}{r^2 - \mathbf{b}^2} + \frac{\mathbf{c}^2 z^2}{r^2 - \mathbf{c}^2} = 0. \quad . \quad . \quad \textbf{(32·371)}$$

§ 32·4. RECIPROCAL RELATIONSHIPS BETWEEN WAVE VELOCITY AND RAY VELOCITY

We may obviously express Fresnel's equation **(25·48)** in the form :

$$\frac{\alpha^2 \xi^2}{\omega^2 - \alpha^2} + \frac{\beta^2 \eta^2}{\omega^2 - \beta^2} + \frac{\gamma^2 \varsigma^2}{\omega^2 - \gamma^2} = 0, \quad . \quad . \quad \textbf{(32·4)}$$

where α, β and γ mean the reciprocals of $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$ respectively ; ξ, η and ς are ωl, ωm and ωn respectively, ω being the reciprocal of v. Similarly if λ, μ and ν are the direction cosines of a ray, and if ρ means the reciprocal of r, we may derive from **(32·371)** the following equation :

$$\frac{\lambda^2}{\rho^2 - \alpha^2} + \frac{\mu^2}{\rho^2 - \beta^2} + \frac{\nu^2}{\rho^2 - \gamma^2} = 0, \quad . \quad . \quad \textbf{(32·41)}$$

which is the exact analogue of Fresnel's equation, and which shows that the reciprocal of the ray velocity is the same function

of the direction and of the constants α^2, β^2 and γ^2 that the wave velocity is of the direction and the constants a^2, b^2 and c^2.

Equation (32·4) is clearly the equation of a surface, identical in shape with the wave surface. Each of its *radii vectores* represents the reciprocal of v^2 (the square of the wave velocity) along its direction; while each radius vector of the wave surface represents the square of the ray velocity along its direction.

And since by (32·21)

$$r \cos \phi = v,$$

we see that ω and ρ are connected by

$$\omega \cos \phi = \rho. \qquad . \qquad . \qquad . \qquad . \qquad (32 \cdot 42)$$

It follows that we may regard the surface (32·4) as the envelope of all the planes

$$\lambda \xi + \mu \eta + \nu \varsigma = \rho. \qquad . \qquad . \qquad . \qquad (32 \cdot 43)$$

Each section of either of the surfaces, (32·371) or (32·4), by the principal planes, $X = 0$, $Y = 0$, or $Z = 0$, consists of a circle and an ellipse. In the case of the section by $Y = 0$ the circle (the radius of which is equal to β or to b as the case may be) and the ellipse cut one another, the points of intersection, a, b, c, d, in Fig. 32·4, determining the directions of the optic axes in the case of (32·4) and those of the axes of single ray velocity in the case of the wave surface. Each *radius vector* drawn in any other direction cuts the surface (32·4) twice in accordance with the fact that in a given direction (other than that of an optic axis) there are *two* wave velocities, and each *radius vector* of the wave surface cuts it, in general, twice : in a given direction there are two ray velocities. Further, any direction of plane wave propagation is associated, in general, with two rays of different velocities and directions represented by the *radii vectores* from O to the points of contact of the wave surface and the corresponding planes. A single incident ray therefore gives rise to *two* rays in a crystal (double refraction). In the case of a uniaxial crystal one of the rays obeys the ordinary laws of refraction, i.e. the laws characteristic of an isotropic medium, since the corresponding part of the wave surface (cf. 32·15) is a sphere of radius a or c as the case may be.

Fig. 32·4

§ 32·5. CONICAL REFRACTION

Each of the points a, b, c, d (Fig. 32·4) is the apex of a conical depression in the outer wave surface. At such a point there is an infinite number of tangent planes. Consequently a direction of single wave velocity (optic axis) is associated with an infinite number of rays. The directions of these are represented by the normals to the tangent planes at one of the points a, b, c, d of the surface **(32·4)**. A tangent plane to **(32·371)** representing a single wave velocity is described by the equation

$$l_0 x + n_0 y = \mathbf{b},$$

where l_0 and n_0 have the values **(32·14)**, and it touches the wave surface **(32·371)** in a closed loop, which, we shall see, is a circle. The *radii vectores* from O to the points on this circle represent the ray velocities in direction and magnitude. The associated phenomenon is called **internal conical refraction.**

If we multiply **(32·36)** by l, m and n respectively and add, we obtain

$$\frac{xl}{r^2 - \mathbf{a}^2} + \frac{ym}{r^2 - \mathbf{b}^2} + \frac{zn}{r^2 - \mathbf{c}^2} = 0.$$

On substituting for l, m and n the values **(32·14)** appropriate for an optic axis we get

$$\frac{xl_0}{r^2 - \mathbf{a}^2} + \frac{zn_0}{r^2 - \mathbf{c}^2} = 0,$$

or

$$r^2(xl_0 + zn_0) = \mathbf{c}^2 l_0 x + \mathbf{a}^2 n_0 z.$$

But $xl_0 + zn_0 = \mathbf{b}$; therefore

$$r^2 = \frac{\mathbf{c}^2 l_0 x}{\mathbf{b}} + \frac{\mathbf{a}^2 n_0 z}{\mathbf{b}}. \qquad \qquad \textbf{(32·5)}$$

This is the equation of a sphere through the origin, and since the points of contact are on a plane, namely on $l_0 x + n_0 z = \mathbf{b}$, as well as on the sphere **(32·5)**, they lie of a circle. It is easy to see that the *radii vectores*, from the origin to the circle of contact, form the cone

$$\mathbf{b}^2 r^2 = (\mathbf{c}^2 l_0 x + \mathbf{a}^2 n_0 z)(l_0 x + n_0 z). \quad . \quad \textbf{(32·51)}$$

The reciprocal case of **external conical refraction** can be dealt with in precisely the same way.

§ 32·6. The Direction of the Displacement

It follows from (25·47) that

$$\alpha m - \beta l = G'lm \frac{\mathbf{b}^2 - \mathbf{a}^2}{(\mathbf{a}^2 - v^2)(\mathbf{b}^2 - v^2)},$$

when α, β and γ are the direction cosines of the displacement in a plane wave travelling in the direction (l, m, n), and G' is defined by (25·462). Now by (32·36) we have

$$\nu = \frac{vn(r^2 - \mathbf{c}^2)}{r(v^2 - \mathbf{c}^2)},$$

since $z = r\nu$; hence

$$\nu(\alpha m - \beta l) = \frac{G'vlmn\{(r^2 - \mathbf{c}^2)(\mathbf{b}^2 - \mathbf{a}^2)\}}{r(v^2 - \mathbf{a}^2)(v^2 - \mathbf{b}^2)(v^2 - \mathbf{c}^2)}.$$

We can, in a similar way, or merely by a cyclic interchange of the symbols in the last equation, obtain expressions for $\lambda(\beta n - \gamma m)$ and $\mu(\gamma l - \alpha n)$. On adding we find that the sum

$$\lambda(\beta n - \gamma m) + \mu(\gamma l - \alpha n) + \nu(\alpha m - \beta l) = 0,$$

or

$$\begin{vmatrix} \alpha, & \beta, & \gamma \\ l, & m, & n \\ \lambda, & \mu, & \nu \end{vmatrix} = 0. \quad . \quad . \quad . \quad . \quad (32 \cdot 6)$$

According to (2·16) this is the condition that the three vectors (α, β, γ), (l, m, n) and (λ, μ, ν), if represented by lines drawn from the origin of co-ordinates, are in the same plane. Thus the normal to a tangent plane of the wave surface, the *radius vector* to the point of contact and the direction of displacement in the corresponding plane wave are co-planar.

§ 32·7. Displacement in Plane Waves travelling in THE SAME Direction

In a given direction (l, m, n) two plane waves are propagated with velocities which we may call v_1 and v_2. Let us represent the direction cosines of the displacements in these waves by $(\alpha_1, \beta_1, \gamma_1)$ and $(\alpha_2, \beta_2, \gamma_2)$ respectively. From (25·47) we have

$$\alpha_1\alpha_2 = \frac{G_1'\ G_2'\ l^2}{(v_1{}^2 - \mathbf{a}^2)(v_2{}^2 - \mathbf{a}^2)}.$$

The cosine of the angle, ε, between these two directions is

$$\cos \varepsilon = \alpha_1\alpha_2 + \beta_1\beta_2 + \gamma_1\gamma_2 = G_1'G_2' \times$$

$$\left\{ \frac{l^2}{(v_1{}^2-\mathbf{a}^2)\,(v_2{}^2-\mathbf{a}^2)} + \frac{m^2}{(v_1{}^2-\mathbf{b}^2)\,(v_2{}^2-\mathbf{b}^2)} + \frac{n^2}{(v_1{}^2-\mathbf{c}^2)\,(v_2{}^2-\mathbf{c}^2)} \right\},$$

or

$$\cos \varepsilon = \frac{G_1'G_2'}{v_2{}^2 - v_1{}^2} \left\{ \left(\frac{l^2}{v_1{}^2 - \mathbf{a}^2} + \frac{m^2}{v_1{}^2 - \mathbf{b}^2} + \frac{n^2}{v_1{}^2 - \mathbf{c}^2} \right) \right.$$

$$\left. - \left(\frac{l^2}{v_2{}^2-\mathbf{a}^2} + \frac{m^2}{v_2{}^2-\mathbf{b}^2} + \frac{n^2}{v_2{}^2-\mathbf{c}^2} \right) \right\}.$$

So long therefore as $v_2{}^2 \neq v_1{}^2$, i.e. for any direction other than that of an optic axis, $\cos \varepsilon = 0$ by (25·48). The two directions of displacement are therefore at right angles to one another. In the special case of single wave velocity, i.e. when (l, m, n) represents an optic axis, every direction perpendicular to (l, m, n) is in general a possible direction of displacement ; since in this case there is an infinite number of ray directions, each of which, by § 32·6, determines a direction of displacement.

§ 32·8. Direction of the Displacements associated with a given Ray Direction

According to (32·36)

$$\frac{x}{v(r^2 - \mathbf{a}^2)} = \frac{l}{v^2 - \mathbf{a}^2},$$

and since $x = r\lambda$,

$$\frac{r\lambda}{v(r^2 - \mathbf{a}^2)} = -\frac{\alpha}{G'}, \quad . \quad . \quad . \quad . \quad (32\text{·}8)$$

by (25·47).

In general each ray direction is associated with two directions of plane wave propagation, and consequently with two directions of displacement, $(\alpha_1, \beta_1, \gamma_1)$ and $(\alpha_2, \beta_2, \gamma_2)$. If we call the angle between these, w,

$$\cos w = \alpha_1 \alpha_2 + \beta_1 \beta_2 + \gamma_1 \gamma_2,$$

and therefore, by (32·8)

$$\cos w = \frac{r_1 r_2 G_1' G_2'}{v_1 v_2} \left\{ \frac{\lambda^2}{(r_1{}^2 - \mathbf{a}^2)\,(r_2{}^2 - \mathbf{a}^2)} + \frac{\mu^2}{(r_1{}^2 - \mathbf{b}^2)\,(r_2{}^2 - \mathbf{b}^2)} \right.$$

$$\left. + \frac{\nu^2}{(r_1{}^2 - \mathbf{c}^2)\,(r_2{}^2 - \mathbf{c}^2)} \right\}.$$

This may be expressed in the form :

$$\cos w = \frac{r_1 r_2 G_1{}' G_2{}'}{v_1 v_2 (r_2{}^2 - r_1{}^2)} \left\{ \left(\frac{\lambda^2}{r_1{}^2 - \mathbf{a}^2} + \frac{\mu^2}{r_1{}^2 - \mathbf{b}^2} + \frac{\nu^2}{r_1{}^2 - \mathbf{c}^2} \right) \right.$$

$$\left. - \left(\frac{\lambda^2}{r_2{}^2 - \mathbf{a}^2} + \frac{\mu^2}{r_2{}^2 - \mathbf{b}^2} + \frac{\nu^2}{r_2{}^2 - \mathbf{c}^2} \right) \right\},$$

and therefore

$$\cos w = \frac{r_1 r_2 G_1{}' G_2{}'}{v_1 v_2 (r_2{}^2 - r_1{}^2)} \left\{ \frac{1}{r_1{}^2} - \frac{1}{r_2{}^2} \right\}$$

$$= \left(\frac{G_1{}'}{v_1 r_1} \right) \left(\frac{G_2{}'}{v_2 r_2} \right). \quad . \quad . \quad . \quad . \quad (32 \cdot 81)$$

Now we have from **(25·47)** by squaring and adding :

$$\frac{1}{G'^2} = \frac{l^2}{(v^2 - \mathbf{a}^2)^2} + \frac{m^2}{(v^2 - \mathbf{b}^2)^2} + \frac{n^2}{(v^2 - \mathbf{c}^2)^2},$$

and therefore $G'^2 = v^2 (r^2 - v^2)$, by **(32·35)**. On substituting this expression for G' in **(32·81)** we obtain

$$\cos w = \left(1 - \frac{v_1{}^2}{r_1{}^2} \right)^{1/2} \left(1 - \frac{v_2{}^2}{r_2{}^2} \right)^{1/2},$$

and finally

$$\cos w = \sin \phi_1 . \sin \phi_2, \quad . \quad . \quad . \quad (32 \cdot 82)$$

ϕ_1 and ϕ_2 being the angles between the direction of the ray and those of the normals to the respective tangent planes to the wave surface.

In the special cases where the ray coincides with one or both of the directions of the associated wave velocities, ϕ_1 or ϕ_2 or both will be zero and the angle, w, in consequence a right angle.

§ 32·9. Rotation of the Plane of Polarization

Certain media, among them both crystalline media, e.g. quartz, and isotropic media, e.g. cane sugar (dextrose) and laevulose, cause the plane of polarization of a beam of plane polarized light to rotate. If, for instance, the plane of polarization of a horizontal (parallel) beam, before its entrance into the active medium, were horizontal, it would be inclined, in the emergent beam, at an angle to the horizontal proportional to the distance traversed in the medium. For the rest the amount of the rotation is determined by the nature of the medium ; for example, by the concentration in the case of a sugar solution. The immediate cause of this rotation, as experiment shows,

is the splitting of the incident beam into two **circularly polarized** beams which travel with slightly different velocities. A right-handed circularly polarized beam, i.e. one in which circular displacement is related to the direction in which the wave travels in the same way as the rotation of an ordinary screw is related to the direction in which it travels, can be described by the equations :

$$s_x = A \cos \omega\left(t - \frac{z}{v} + \frac{\tau}{4} \right), \\ s_y = A \cos \omega\left(t - \frac{z}{v} \right). \Bigg\} \quad . \quad . \quad . \quad (32 \cdot 9)$$

The wave is travelling in the Z direction with the (phase) velocity, v ; A is the amplitude ; s_x and s_y are the X and Y components of the displacement in the wave and τ is the period. A beam of the same amplitude and frequency, but circularly polarized in the opposite sense and travelling in the same direction with a slightly different velocity, v', will be represented by

$$s_x' = A \cos \omega\left(t - \frac{z}{v'} - \frac{\tau}{4} \right), \\ s_y' = A \cos \omega\left(t - \frac{z}{v'} \right). \Bigg\} \quad . \quad (32 \cdot 901)$$

Let us write

$$\omega\left(t - \frac{z}{v} \right) = \phi, \\ \omega\left(t - \frac{z}{v'} \right) = \phi'. \Bigg\} \quad . \quad . \quad . \quad (32 \cdot 91)$$

These two waves constitute, in virtue of the principle of super-position, the single plane wave

$$S_x = s_x + s_x' = 2A \cos \frac{\phi + \phi'}{2} \cos \frac{\phi - \phi' + \pi}{2}, \\ S_y = s_y + s_y' = 2A \cos \frac{\phi + \phi'}{2} \cos \frac{\phi - \phi'}{2}. \Bigg\} . \quad . \quad (32 \cdot 92)$$

If we define an angle, ε, by

$$\varepsilon \equiv (\phi - \phi' + \pi)/2,$$

or $\qquad\qquad \varepsilon \equiv \omega\left(\frac{1}{v'} - \frac{1}{v} \right) z/2 + \pi/2, \quad . \quad . \quad . \quad (32 \cdot 921)$

we may write equations **(32·92)** in the form :

$$S_x = 2A \cos \varepsilon \cos \omega \left\{ t - \tfrac{1}{2}\left(\frac{1}{v} + \frac{1}{v'}\right)z \right\},$$

$$S_y = 2A \sin \varepsilon \cos \omega \left\{ t - \tfrac{1}{2}\left(\frac{1}{v} + \frac{1}{v'}\right)z \right\}. \qquad (32\cdot93)$$

It is seen that the resulting wave has an amplitude, $2A$, and the resulting displacement is along the straight line which makes the angle, ε, with the X axis. This angle varies with z, i.e. with the distance traversed in the active medium, in the way shown by **(32·921)**. The velocity of the wave is, $2vv'(v + v')$, in accordance with **(32·93)**.

The two oppositely circularly polarized beams may be due, as in the case of sugar or quartz, to peculiarities in molecular structure ; but they can also be produced by a magnetic field. For more complete information the reader is referred to Drude's *Lehrbuch* ; we shall only mention here one interesting difference between the two types of rotation. When the rotation is due solely to the peculiar character of the molecular structure, it is reversed and annulled when the beam is reflected and caused to re-traverse the medium in the opposite direction ; when the rotation is due to an external magnetic field in the direction of propagation, it is continued, and consequently doubled, when the beam is caused to re-traverse the medium in the opposite direction.

BIBLIOGRAPHY

P. Drude : Lehrbuch.
Clemens Schäfer : Einführung. (Band III.)
M. Planck : Introduction to Theoretical Physics. (Macmillan.) Volume IV is devoted to the Theory of Light.

CHAPTER XV

GENERAL QUESTIONS CONCERNING THE PROPAGATION OF LIGHT

§ 33.   Absolute Measurements of the Velocity of Light give the Group Velocity

ANY method of measurement in which the distance traversed and the time required are observed, must give the group velocity.   The observer has no means of taking note of a wave crest as it travels through some measured distance, and even if he had such a means it would be of doubtful value for obtaining the phase velocity, because of the limited range through which an individual crest travels between the instants of its creation and vanishing.   Quite apart from the fact that the phenomenon (e.g. the eclipsing of Jupiter's satellites) or the apparatus used (Fizeau's toothed wheel or Foucault's rotating mirror) breaks up the light into trains of waves, i.e. into groups, the very processes of emission in the source of light do this ;   and when the phase and group velocities differ, crests and troughs will be continually produced at one limit of the group and destroyed at the other.   The observer can only take note of the group.

In empty space,[1] or in any non-dispersive medium, the two velocities are identical and the distinction between them is not of *practical* importance.   This follows from the expressions obtained for the two velocities in § 9·3.   The group velocity is expressed by

$$v = d(1/\tau)/d(1/\lambda),$$

or

$$v = \frac{\lambda^2}{\tau^2}\frac{d\tau}{d\lambda}. \qquad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad (33)$$

[1] We use the term ' empty space ' for a region which has been exhausted by a pump, or for a region which is empty in the same sense ; so that, if not actually empty, it is occupied solely by the hypothetical aether.

The phase velocity, u, is of course represented by

$$u = \lambda/\tau, \quad . \quad . \quad . \quad . \quad . \quad \textbf{(33·01)}$$

and when there is no dispersion,

$$u = d\lambda/d\tau,$$

so that in this case

$$\frac{d\lambda}{d\tau} = \frac{\lambda}{\tau}.$$

Substituting this in **(33)** we find

$$v = \lambda/\tau = u.$$

Bradley's method, based on the measurement of the angle of aberration of the light from a star, has features which place it in a category by itself, and cannot be completely discussed until we deal with the theory of relativity. But if the intervening medium were dispersive the angle of aberration would yield the group velocity.

§ 33·1. The Aether

The conception of a luminiferous medium, or aether, occupying ' empty ' space and penetrating materials as well, has played a great and useful part in physical theories in the past, and very naturally experiments of various types have been devised and carried out to observe it, as we might say, and to measure its velocity relative to material bodies. The most significant of these was Michelson's experiment **(§§ 31·3 and 33·4)**, the purpose of which was to detect and measure suspected slight variations of the velocity with direction. The expected result was of the order of magnitude of v^2/c^2, compared with unity (a second order small change) v being the velocity of the earth or of the observer relative to the aether. The result was negative.

Another type of experiment which yields a positive first order result was carried out by Fizeau in 1851, and confirmed later by Michelson and still more recently by Zeeman. It was originally devised to test a prediction made by Fresnel more than thirty years earlier. The significance of these experiments will be discussed in the following sections.

§ 33·2. Fresnel's Convection Coefficient and Fizeau's Experiment

About the year 1818 Arago demonstrated experimentally that the laws of refraction of the light from the stars were un-

affected by the motion of the earth. It is instructive to examine this result in the light of Fermat's principle (**§ 28·4**). Imagine a ray of light to start from some point A (x_1, y_1, z_1) and, after travelling some distance in empty space, to be refracted at a plane surface, at the point (x, y, z), and finally to reach some point B (x_2, y_2, z_2), as in Fig. 28·5. The plane surface is that of a transparent material medium travelling with a given velocity, v.[1] The laws of refraction are, by hypothesis, not affected by this motion. Let us for convenience assume B is a definite point in the material medium and by hypothesis moves with it. We shall for convenience imagine A also to move with the velocity v.

The axes of co-ordinates are fixed relatively to the moving medium. If T be the time required by a wave crest to travel from A to B, the path chosen by the light is such that

$$\delta T = 0. \quad . \quad . \quad . \quad . \quad . \quad \textbf{(33·2)}$$

In order to obtain an expression for T we must use the phase velocity *relatively to the moving system*. If the velocity, v, were in exactly the same direction as that in which the light is travelling, we should have, for u_1, the relative velocity in empty space,

$$u_1 = c - v, \quad . \quad . \quad . \quad . \quad \textbf{(33·21)}$$

and for u_2, the relative velocity in the material medium,

$$u_2 = c' - v,$$

where c' is the velocity in the material. Now

$$c' = c/\mu + v'$$

if v' means the increment of velocity due to the motion, or, as we might describe it, the convective effect of the motion of the material medium, and if μ means the refractive index of the material. Thus we should have for u_2

$$u_2 = c/\mu + (v' - v),$$

or, if we define k by

$$k \equiv v'/v, \quad . \quad . \quad . \quad . \quad \textbf{(33·211)}$$
$$u_2 = c/\mu - (1 - k)v. \quad . \quad . \quad \textbf{(33·212)}$$

The fraction, k, is known as **Fresnel's convection coefficient.**

In the general case when v is not necessarily in the same direction as that in which the light is travelling, we have instead of **(33·21)** and **(33·212)**,

$$u_1 = c - v \cos \theta_1, \quad . \quad . \quad . \quad . \quad \textbf{(33·22)}$$

[1] Relative to the aether outside the material medium.

and
$$u_2 = c/u - (1 - k)v \cos \theta_2, \quad . \quad . \quad (33\cdot221)$$

if we ignore second order small quantities (Fig. 33·2 (a)). Hence (33·2) becomes

$$\delta \left\{ \frac{r_1}{c - v \cos \theta_1} + \frac{r_2}{c/\mu - (1 - k)v \cos \theta_2} \right\} = 0,$$
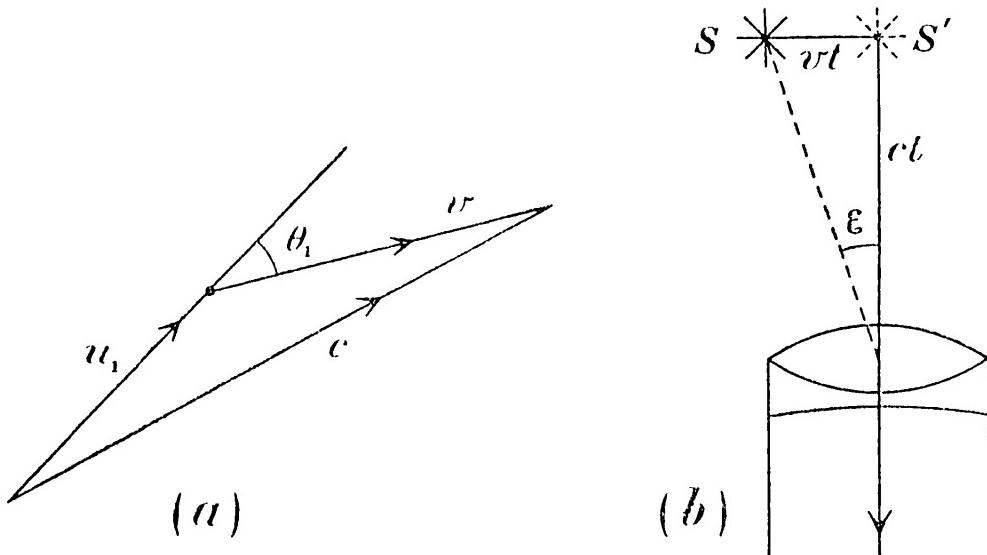
FIG. 33·2

or, still ignoring second order small quantities,

$$\delta \left\{ r_1 \left(1 + \frac{v}{c} \cos \theta_1 \right) + \mu r_2 \left(1 + \frac{\mu(1 - k)v}{c} \cos \theta_2 \right) \right\} = 0 ;$$

in which r_1 and r_2 have the same meanings as in § **28·5** and Fig. 28·5. The last equation is, of course, valid when $v = 0$, and therefore

$$\delta(r_1 + \mu r_2) = 0.$$

Consequently

$$\delta \left\{ r_1 v \cos \theta_1 + \mu^2 r_2 (1 - k)v \cos \theta_2 \right\} = 0. \quad . \quad (33\cdot23)$$

Now

$$r_1 v \cos \theta_1 = (\mathbf{r_1}, \mathbf{v}) = (x_1 - x)v_x + (y_1 - y)v_y + (z_1 - z)v_z,$$

and

$$r_2 v \cos \theta_2 = (\mathbf{r_2}, \mathbf{v}) = (x - x_2)v_x + (y - y_2)v_y + (z - z_2)v_z.$$

Therefore (33·23) becomes

$$\{\mu^2(1 - k) - 1\}(v_x \delta x + v_y \delta y + v_z \delta z) = 0. \quad . \quad (33\cdot24)$$

If
$$Ax + By + Cz + D = 0$$

be the equation of the refracting surface,

$$\delta x = -\frac{B}{A}\delta y - \frac{C}{A}\delta z,$$

and (33·24) may be written :

$$\{\mu^2(1 - k) - 1\}\left\{\left(v_y - \frac{B}{A}v_x\right)\delta y + \left(v_z - \frac{C}{A}v_x\right)\delta z\right\} = 0.$$

The right-hand factor in this equation is quite arbitrary, and it follows that
$$\mu^2(1 - k) - 1 = 0,$$
or

$$k = \left(1 - \frac{1}{\mu^2}\right). \quad . \quad . \quad . \quad (33\cdot25)$$

It follows, therefore, that when the medium is moving with a velocity, v, in the same direction as the light, the phase velocity is increased from c/μ to

$$c/\mu + (1 - 1/\mu^2)v. \quad . \quad . \quad . \quad (33\cdot26)$$

This formula must be carefully interpreted ; μ means the refractive index of the light that is in the moving medium. In consequence of the motion, however, the period, τ, of the light at a point fixed relatively to the medium is slightly different from what it would be if the medium were at rest. If, therefore, μ were used to mean the refractive index of the same light in the quiescent material, we should have to make a slight change in the term c/μ. We need not trouble about the term $(1 - 1/\mu^2)v$ since the modification of this would only produce a second order small change. The change in the period, $d\tau$, is easily seen to be

$$d\tau = \frac{\mu v}{c}\tau,$$

if we neglect second order small quantities. This is simply the change in period known as the *Doppler effect*. Hence we must replace c/μ by

$$\frac{c}{\mu}\left(1 - \frac{v\tau}{c}\frac{d\mu}{d\tau}\right)$$

and (33·26) must be replaced by

$$\frac{c}{\mu} + \left\{1 - \frac{1}{\mu^2} - \frac{\tau}{\mu}\frac{d\mu}{d\tau}\right\}v, \quad . \quad . \quad . \quad (33\cdot27)$$

in which μ now means the refractive index of the same light in the quiescent medium.

Fizeau verified (33·26) experimentally by producing interference fringes with coherent beams of light which travelled through a water column in opposite senses. When the water was set in motion, the velocity of one beam was increased and that of the other diminished with a consequent measurable displacement of the fringes from which the velocity changes could be calculated.

§ 33·3. The Convection of Light and Stellar Aberration

The subject of the last section has a bearing on the phenomenon of aberration. For simplicity let us take the case where the direction of the star is at right angles to that of the motion of the observer and his telescope. The simple or naïve view of aberration leads to the formula :

$$v/c = \tan \varepsilon, \quad . \quad . \quad . \quad . \quad . \quad (33\cdot3)$$

on the assumption that the optical medium, both without and within the telescope, is the free aether which is at ' rest '. The telescope is moving through it with the velocity, v ; while c is the velocity of light and ε the angle of aberration. Let S' (Fig. 33·2 (b)) be the position of the star relative to the telescope at an instant when a group of waves is emitted. The further sequence of events is, if we neglect second order small quantities, precisely that to be expected if the star S' and telescope are at rest relatively to one another. The formulae (32·26) and (32·27) are based on this assumption. If therefore at this instant the telescope were directed to the star, the central ray in the group would travel down the axis of the telescope. This would be true even if the telescope were filled with water or any other transparent medium. If t be the time taken by the light to reach the objective, the distance travelled by the light is ct. The star, however, is in fact at rest at S, and therefore separated from S' by the distance $SS' = vt$. Hence

$$vt/ct = v/c = \tan \varepsilon.$$

In other words, the observed angle of aberration is independent of the nature of the optical medium filling the telescope. This deduction was verified by the astronomer Airy, who had a telescope set up at Greenwich, the tube of which was filled with water. The formula (33·3) must be modified to

$$v/c = \sin \varepsilon, \quad . \quad . \quad . \quad . \quad . \quad (33\cdot301)$$

as we shall see later, if we adopt the principles of the theory of relativity.

§ 33·4. MICHELSON'S EXPERIMENT

Let us suppose the two optical paths OA and OB (Fig. 31·2) from the point O in Michelson's interfero-meter, where the original beam is divided, to the mirrors A and B, to be equal. Imagine the ap-paratus to be moving with the velocity, v, relatively to the aether in the direction of OA (Fig. 33·4), and let us find expressions for the times T_1 and T_2 required by the light to travel

FIG. 33·4

from O to A and back, and from O to B and back, respectively. Obviously
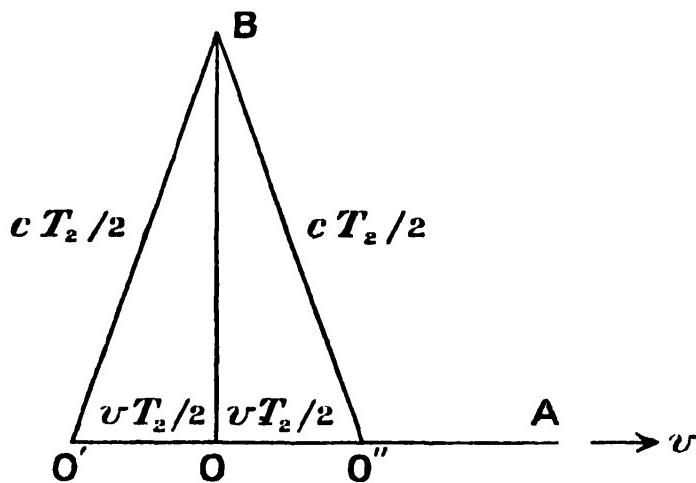
$$T_1 = L\left\{\frac{1}{c - v} + \frac{1}{c + v}\right\}$$

or

$$T_1 = \gamma^2 \frac{2L}{c}, \quad . \quad . \quad . \quad . \quad . \quad . \quad \textbf{(33·4)}$$

where L is the length of the arm OA and γ means $(1 - v^2/c^2)^{-1/2}$. The light reflected from B starts from O when it is in the position O' (Fig. 33·4) and returns to it again when it has reached O". Therefore

$$L^2 = \frac{(c^2 - v^2)T_2^2}{4},$$

and

$$T_2 = \gamma \frac{2L}{c}. \quad . \quad . \quad . \quad . \quad . \quad \textbf{(33·41)}$$

The experiment indicated that

$$T_1 = T_2.$$

One obvious way of accounting for this result is to assume $v = 0$, i.e. that the aether is moving with the earth. This assumption, however, makes it exceedingly difficult to explain stellar aberration.

20*

We have to fall back on the FitzGerald-Lorentz contraction hypothesis (cf. **§ 28**) that the dimensions of bodies are reduced along the line of motion in the ratio $\gamma : 1$.   So that **(33·4)** becomes

$$T_1 = \gamma^2 \frac{2L'}{c},$$

where

$$\gamma L' = L,$$

and T_1 and T_2 are now equal.

The Lorentz-FitzGerald hypothesis brings us once again into immediate contact with the theory of relativity, in the light of which the further discussion of the subjects of this chapter must be continued.

BIBLIOGRAPHY

J. LARMOR : Aether and Matter. (Cambridge.)

T. H. HAVELOCK : The Propagation of Disturbances in Dispersive Media (Chapter II). (Cambridge Tracts.)

INDEX OF SUBJECTS

INDEX OF NAMES